2

AAMRL-TR-90-007

# EVALUATION OF THE CRITERION TASK SET — PART I CTS PERFORMANCE AND SWAT DATA — BASELINE CONDITIONS (U)

Robert E. Schlegel, Ph.D.
Kirby Gilliland, Ph.D.

THE UNIVERSITY OF OKLAHOMA

JANUARY 1990

FINAL REPORT FOR MAY 1988 - MAY 1989

Approved for public release; distribution is unlimited.

HARRY G. ARMSTRONG AEROSPACE MEDICAL RESEARCH LABORATORY
HUMAN SYSTEMS DIVISION
AIR FORCE SYSTEMS COMMAND
WRIGHT-PATTERSON AIR FORCE BASE, OHIO 45433-6573

## NOTICES

When US Government drawings, specifications, or other data are used for any purpose other than a definitely related Government procurement operation, the Government thereby incurs no responsibility nor any obligation whatsoever, and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication or otherwise, as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

Please do not request copies of this report from the Armstrong Aerospace Medical Research Laboratory. Additional copies may be purchased from:

National Technical Information Service
5285 Port Royal Road
Springfield, Virginia 22161

Federal Government agencies and their contractors registered with the Defense Technical Information Center should direct requests for copies of this report to:

Defense Technical Information Center
Cameron Station
Alexandria, Virginia 22314

## TECHNICAL REVIEW AND APPROVAL

AAMRL-TR-90-007

This report has been reviewed by the Office of Public Affairs (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS, it will be available to the general public, including foreign nations.

The voluntary informed consent of the subjects used in this research was obtained as required by Air Force Regulation 169-3.

This technical report has been reviewed and is approved for publication.

**FOR THE COMMANDER**

Charles Bates

CHARLES BATES, JR.
Director, Human Engineering Division
Armstrong Aerospace Medical Research Laboratory

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

## REPORT DOCUMENTATION PAGE

*Form Approved OMB No 0704-0188*

1a REPORT SECURITY CLASSIFICATION
Unclassified

1b RESTRICTIVE MARKINGS

2a SECURITY CLASSIFICATION AUTHORITY

2b DECLASSIFICATION / DOWNGRADING SCHEDULE

3 DISTRIBUTION / AVAILABILITY OF REPORT
Approved for public release; distribution is unlimited

4 PERFORMING ORGANIZATION REPORT NUMBER(S)

5 MONITORING ORGANIZATION REPORT NUMBER(S)
AAMRL-TR-90-007

6a NAME OF PERFORMING ORGANIZATION
The University of Oklahoma

6b OFFICE SYMBOL *(If applicable)*

7a NAME OF MONITORING ORGANIZATION
Harry G. Armstrong Aerospace Medical Research Laboratory

6c ADDRESS *(City, State, and ZIP Code)*
1000 Asp, Room 314
Norman, Oklahoma 73019

7b ADDRESS *(City, State, and ZIP Code)*
Wright-Patterson Air Force Base,
Ohio 45433-6573

8a NAME OF FUNDING / SPONSORING ORGANIZATION

8b OFFICE SYMBOL *(If applicable)*

9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
F33615-85-D-0540

8c ADDRESS *(City, State, and ZIP Code)*

10 SOURCE OF FUNDING NUMBERS

| PROGRAM ELEMENT NO | PROJECT NO | TASK NO | WORK UNIT ACCESSION NO |
|---|---|---|---|
| 62202F | 7184 | 00 | 13 |

11 TITLE *(Include Security Classification)*
Evaluation of the Criterion Task Set - Part I, CTS Performance and SWAT Data - Baseline Conditions (U)

12 PERSONAL AUTHOR(S)
Schlegel, Robert E. PhD, and Gilliland, Kirby Ph.D.

13a TYPE OF REPORT
Final

13b TIME COVERED
FROM 2 May88 TO 2 May89

14 DATE OF REPORT *(Year, Month, Day)*
1990 January

15 PAGE COUNT
179

16 SUPPLEMENTARY NOTATION
*Subcontract to Southeastern Center for Electrical Engineering Education, Central Florida Facility, 1101 Massachusetts Avenue, St. Cloud, Florida 32769

17 COSATI CODES

| FIELD | GROUP | SUB-GROUP |
|---|---|---|
| 05 | 08 | |
| 23 | 02 | |

18 SUBJECT TERMS *(Continue on reverse if necessary and identify by block number)*
Human Performance,
Task Battery
Cognition

19 ABSTRACT *(Continue on reverse if necessary and identify by block number)*

This report summarizes the development and analysis of a comprehensive standardization data base for the USAF Criterion Task Set (CTS). The CTS is a collection of standardized loading tasks developed as a mental workload metric evaluation tool (see AFAMRL-TR-84-071). Performance data, Subjective Workload Assessment Technique (SWAT) data, and individual difference measures were collected and are reported for 123 subjects (95 men, 28 women) for all nine tasks of the CTS Version 1.0. Part I of the Final Report (this document) details the experimental procedures for developing the data base and summarizes the performance data and SWAT ratings with respect to task difficulty levels, learning rates, stability of the measures, gender and SWAT prototype differences, and intertask relationships. As a basis of comparison, the data in this report should be of value to others using the Criterion Task Set to evaluate human information processing performance.

20 DISTRIBUTION / AVAILABILITY OF ABSTRACT
☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS

21 ABSTRACT SECURITY CLASSIFICATION
Unclassified

22a NAME OF RESPONSIBLE INDIVIDUAL
Gary B. Reid

22b TELEPHONE *(Include Area Code)*
(513) 255-8749

22c OFFICE SYMBOL
AAMRL/HEG

*Previous editions are obsolete*

SECURITY CLASSIFICATION OF THIS PAGE

UNCLASSIFIED

# PREFACE

This report documents the work performed at the University of Oklahoma under SCEEE Subcontract HER/86-9 for the Armstrong Aerospace Medical Research Laboratory, Workload and Ergonomics Branch (AAMRL/HEG) under contract F33615-85-D-0514.

As outlined in the task order, a large-scale experimental study was conducted involving the training and testing of 50 additional human subjects on the Criterion Task Set (Version 1.0) under baseline conditions. The complete data base includes performance data, subjective ratings obtained using the Subjective Workload Assessment Technique (SWAT) for each task, and data on subject individual differences (gender, age, personality variables, etc.) for more than 120 subjects. Part I of this report details the procedures for developing the data base and summarizes the performance data and SWAT ratings with respect to task difficulty levels, learning rates, stability of the measures, task intercorrelations, gender differences and personality variables. Part II provides summaries and analyses of performance and SWAT data under noise, sleep deprivation, caffeine and deadline conditions. Part III provides preliminary analyses related to the individual difference variables.

The authors gratefully acknowledge the contributions of graduate research assistants Sharon Dannels, Betina Schlegel and Jen-Gwo Chen and the undergraduate support team in collection, conversion and analysis of the vast amounts of data. They also wish to thank Gary B. Reid, F. Thomas Eggemeier and Clark A. Shingledecker for providing their technical and personal support for the effort.

# SUMMARY

This report summarizes the development and analysis of a comprehensive standardization data base for the USAF Criterion Task Set (CTS). The CTS is a collection of standardized loading tasks developed by the Harry G. Armstrong Aerospace Medical Research Laboratory as a mental workload metric evaluation tool (see AFAMRL-TR-84-071). The effort reported in this document was conducted by the University of Oklahoma. Performance data, Subjective Workload Assessment Technique (SWAT) data, and individual difference measures were collected and are reported for 123 subjects (95 men, 28 women) for all nine tasks of the CTS Version 1.0. Part I of the Final Report (this document) details the experimental procedures for developing the data base and summarizes the performance data and SWAT ratings with respect to task difficulty levels, learning rates, stability of the measures, gender and SWAT prototype differences, and intertask relationships.

For all tasks, the data verified the existence of statistically significant performance and SWAT differences for all three levels of task difficulty. Few gender differences and no SWAT prototype group differences were evident. Cluster analysis provided evidence supporting the CTS design goal of minimum overlap of distinct processing resources across tasks. As a basis of comparison, the data in this report should be of value to others using the Criterion Task Set to evaluate human information processing performance.

# TABLE OF CONTENTS

# LIST OF TABLES

**Table** **Page**

# LIST OF FIGURES

**Figure** **Page**

# EVALUATION OF THE CRITERION TASK SET

## 1.0 INTRODUCTION

In recent years there has been increasing interest in assessing the amount of mental work demanded by a specific operator task. Knowledge of the mental workload associated with an operator's task is crucial in the design of tasks and operational environments. In addition, as the increasing complexity of operator tasks places added demands on mental resources, the effects of factors such as environmental stress, drugs, and individual differences in performance become more important.

Trends in workload research have concentrated on two areas. The first area has been the development of workload theory, that is, models which can predict human information processing/performance capability. The second area of concentration has been the development of specific measurement techniques. Unfortunately, there have been very few attempts to coalesce these theory and task developments into a unified workload assessment technique. However, the U.S. Air Force has developed a mental workload metric evaluation tool, the Criterion Task Set (CTS) which was designed for this purpose (Shingledecker, Acton, and Crabtree, 1983).

The CTS was designed to provide a set of standardized loading tasks to evaluate the relative sensitivity, reliability, and intrusiveness of a variety of available workload measures. In addition, the CTS can be directly used for human performance assessment as well.

The general task of this contract was to design and perform a large-scale data collection and analysis effort to support a comprehensive evaluation of the validity and reliability of the Air Force Criterion Task Set. This volume of the final report summarizes the experimental design and methods used to obtain this data set. It also provides statistical summaries and analyses for the performance and subjective workload data during training and baseline conditions.

### 1.1 Criterion Task Set (CTS)

The CTS is based on a model which represents a synthesis of current information processing theories (Sternberg, 1969; Wickens, 1981). According to these theories, *human mental performance* is dependent on a number of stages, information processing resources, and specific functions. The CTS model hypothesizes three primary stages of processing: perceptual input, central processing, and motor output (Figure 1-1). There

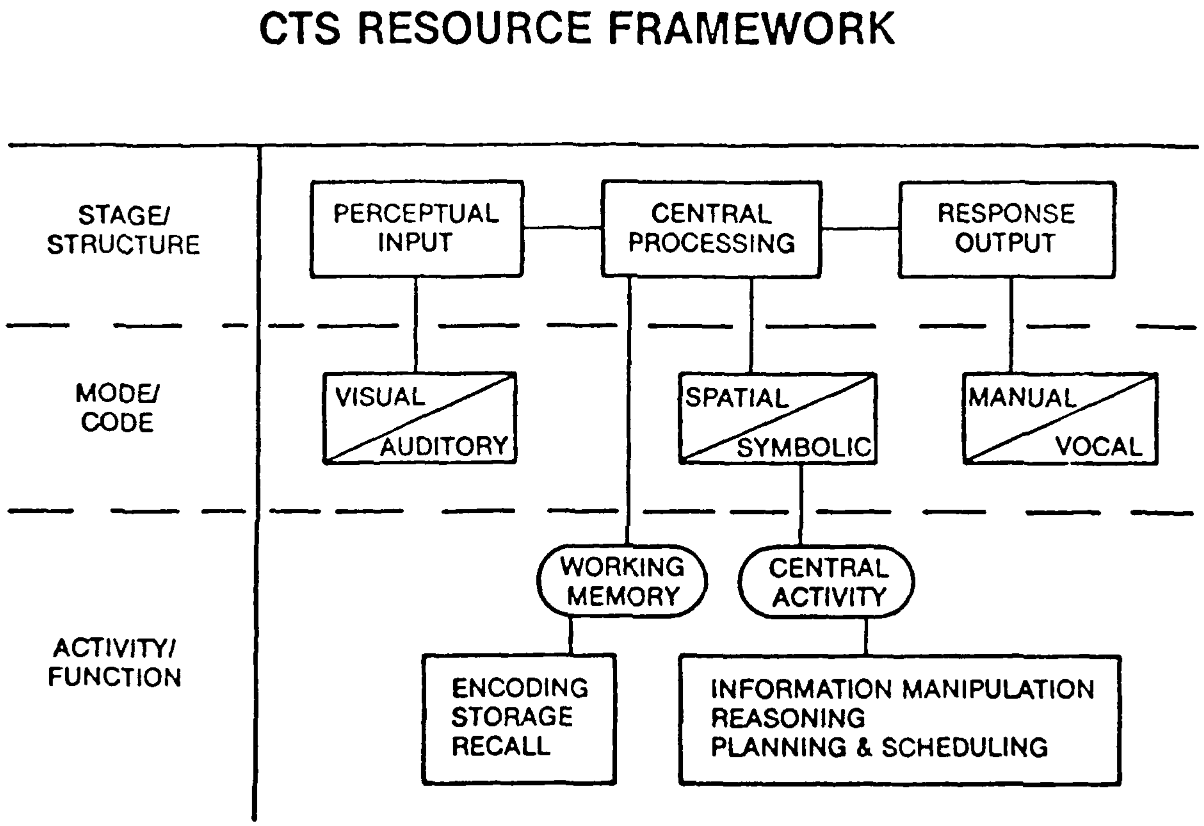


**Figure 1-1. CTS Resource Framework.**
(Shingledecker, 1984)

are specific mental processing resources associated with the input mode (either visual or auditory), the type of coding during central processing (either spatial/imaginal or abstract/symbolic), and the mode of response output (either manual or vocal). Also, the central processing stage is further divided to emphasize memory/recall functions and elementary mental activities such as information manipulation, reasoning, and planning/scheduling.

This model was then used to guide the selection of CTS tasks which, it was assumed, would be representative of the range of human operator performance. This was accomplished by operationally defining each element in the model in terms of the task characteristics associated with the resources required by that element. For example, resources associated with the visual perceptual/input element were defined in terms of the task characteristics of stimulus discriminability and numerosity of display sources. These characteristics would be represented by tasks requiring simple detection as well as monitoring and scanning ability.

Additionally, it was recognized that any task is likely to make demands at all processing stages. Thus, when actually selecting a candidate task for a specific element of the model, such as visual perceptual/input, the loading demands on central processing and motor/output elements were minimized.

A wide range of tasks from the literature on cognitive and psychomotor performance was screened using the definitions as noted above. The screening process resulted in the selection of eleven tasks. Nine of those tasks were selected for inclusion in CTS Version 1.0. Table 1-1 summarizes the characteristics of these tasks presented in alphabetic order.

Initial parametric studies were completed to determine estimates of training time needed for each task, to determine task pacing rates, and to establish standard task loading levels. The standard loading levels were determined through comparison of post-asymptotic performance measures and were corroborated by subjective ratings of task difficulty and complexity (Shingledecker, 1984).

## 1.2 CTS Task Descriptions

The following are brief descriptions of the CTS Version 1.0 tasks. Detailed descriptions and the results of the initial parametric studies are provided by Shingledecker (1984).

## Table 1-1. Summary of CTS Version 1.0 Task Characteristics.

| Task | Code | Stage | Level | Description |
|---|---|---|---|---|
| Continuous Recall | CR | Central | | memory update |
| | | | 1 | 1 digit - next item |
| | | | 2 | 2 digits - 2 items back |
| | | | 3 | 4 digits - 3 items back |
| Grammatical Reasoning | GR | Central | | logic; reasoning |
| | | | 1 | single sentence |
| | | | 2 | two sentences - active/positive |
| | | | 3 | two sentences - active/negative or passive/negative |
| Interval Production | IP | Output | 1 | tapping at 2 taps/sec. |
| Linguistic Processing | LP | Central | | symbol manipulation |
| | | | 1 | physical identity |
| | | | 2 | vowel/consonant |
| | | | 3 | antonyms |
| Mathematical Processing | MP | Central | | math (+ or -) |
| | | | 1 | 1 operation |
| | | | 2 | 2 operations |
| | | | 3 | 3 operations |
| Memory Search | MS | Central | | Sternberg memory test |
| | | | 1 | positive set size 1 |
| | | | 2 | positive set size 4 |
| | | | 3 | positive set size 6 |
| Probability Monitoring | PM | Input | | scanning/detection |
| | | | 1 | 1 meter, 95% bias |
| | | | 2 | 3 meters, 85% bias |
| | | | 3 | 4 meters, 75% bias |
| Spatial Processing | SP | Central | | histogram shapes |
| | | | 1 | 2 bars, 0° |
| | | | 2 | 4 bars, 90° or 270° |
| | | | 3 | 6 bars, 180° |
| Unstable Tracking | UT | Input/Output | | manual response |
| | | | 1 | lambda=1 |
| | | | 2 | lambda=3 |
| | | | 3 | lambda=5 |

**Visual Probability Monitoring (PM).** The subject is required to monitor 1, 3, or 4 simulated meters and determine whether a bias condition is present. A bias condition occurs when a pointer stays on one side of a meter's centerline a higher percentage of time than on the other. The bias time percentages for the 1-, 2-, and 4-dial conditions are 95%, 85%, and 75%, respectively.

**Continuous Recall (CR).** The subject is presented with pairs of 1-, 2-, or 4-digit numbers, one below and the other above a dividing line. The subject must compare the top number with a previously presented number and memorize the bottom number for a later comparison.

**Memory Search (MS).** An initial set of 1, 4, or 6 letters is presented to the subject for memorization. Following this, the subject must identify whether a randomly generated letter is a member of the memorized set.

**Linguistic Processing (LP).** At the low level, the subject decides whether presented pairs of letters are physically identical. At the medium level, the subject decides whether both letters are vowels or consonants. At the high level, the subject determines whether word pairs are antonyms.

**Mathematical Processing (MP).** The subject is required to decide whether the result of a mathematical expression involving 1, 2, or 3 operators (+ or -) is greater than or less than the value 5.

**Spatial Processing (SP).** The subject compares an initial histogram of 2, 4, or 6 bars with a second histogram that has been rotated 0, 180, or 270 degrees.

**Grammatical Reasoning (GR).** Sentences describing the ordinal position of symbols are presented with the symbols positioned below. The subject must determine whether the sentences correctly describe the positioning of the symbols.

**Unstable Tracking (UT).** The subject attempts to maintain the vertical position of a symbolic airplane on a defined line in the center of the screen, using a rotating control knob. The dynamics of the task magnify the control error to prevent stable control.

***Interval Production (IP).*** The subject attempts to create regular timing intervals using a tapping paddle at a rate of 1 to 3 taps per second.

Based on preliminary results, several of the tasks have been modified in the development of CTS Version 2.0. Additional tasks for the CTS are also under development. These tasks will assess planning and scheduling activities characterized by multiattribute decision requirements. These activities are typical in complex system supervision and planning tasks.

## 2.0 OBJECTIVES

The objective of this project was to provide an extensive CTS data base in support of a large-scale evaluation program assessing the reliability and validity of the CTS and its sensitivity to various context (or stressor) variables. At least three important needs were addressed in the performance of this contract. The first need was to initiate development of a comprehensive data base on CTS task performance. The second need was to develop the data base in such a way as to allow multivariate investigations of the response characteristics of the CTS. The third need was to explore the effects of specific context (or intervening) variables on the response characteristics of the CTS.

### 2.1 Comprehensive Data Base

While preliminary standardization data had been collected for the CTS (Shingledecker et al., 1983; Schlegel, 1986), the use of tests within this battery necessitated more comprehensive knowledge of the structure of the CTS. For example, there was a need for basic information regarding the degree of performance variation on the CTS tasks within the context of a large population. There was also a need for additional information regarding the variability in learning rate and training requirements on these tasks. Previously there existed very little data for accurately estimating the reliability (i.e., stability) of CTS performance under standard laboratory conditions.

### 2.2 Multivariate Analyses

The second need was to develop a data set that would allow for more sophisticated multivariate investigations of the CTS. Multivariate analyses would provide information regarding the inter-relationships of the CTS tasks and task levels, as well as their relationships to other types of tests.

For example, a comparison of the responses of subjects during standard laboratory conditions to their responses under stress conditions would help to determine whether the CTS factor structure changes in relation to stress. This is particularly important because some tests may have specific bandwidths of sensitivity to drug-induced or stress-induced performance change. In other words, tasks which assess the limits of performance across workload levels under standard laboratory conditions may be useless at some workload levels during stress or drug conditions. That is, they may show no predictive ability due to "ceiling" or "basement" effects. This type of data would be particularly useful in the planning of dual-task studies and various stress studies.

This type of analysis would also allow for rapid selection of subsets of CTS tests for use in specialized testing applications, such as the development of Tri-Services Drug Screening "Performance Assessment Batteries" (PAB).

Finally, it is important to note that the Air Force has also developed the Subjective Workload Assessment Technique (SWAT -- Reid, 1982; Reid, Shingledecker, and Eggemeier, 1981) which holds considerable promise as a subjective workload metric capable of assessing the impact of context variables. Multivariate and related correlational analyses would aid in understanding the relationship between SWAT ratings and CTS performance.

### 2.3 Context Variable Data

The term "context variable" refers to three general classes of variables that are present in the operational environment and have the potential for influencing operator performance. These classes of variables are:

(1) **Local Environment** (e.g., time, temperature, noise, obscurants to perception, protective clothing),

(2) **Individual Status** (e.g., fatigue, sleep loss, emotional stressors, disease, nutrition, drug use), and

(3) **Long-Term Individual History** (e.g., training, prior experience, gender, age, and important individual differences in such variables as intelligence level, arousal, and task/cognitive ability).

Of those context variables mentioned above, several are of overriding and obvious concern. These include such factors as noise, fatigue, sleep loss, gender, common drug effects, and specific individual difference variables that have been shown to be highly related to performance or cognitive processing ability.

Noise, fatigue, and sleep loss, as well as common drug use, are variables which typically affect performance. Noise is commonly associated with operational environments and has been shown in many cases to be disruptive to performance, especially if loud or distracting. Fatigue and/or sleep loss are major problems in any operational environment and their effects on performance are well documented. While major emphasis could reasonably be placed on chemical defense prophylactic drug effects, there exist considerable data on the disruptive effects of drugs such as caffeine commonly found in the operational environment.

Finally, there are several individual difference variables which are known to relate to the manner in which a person processes information or to the processes related directly to performance capability (e.g., arousal state). There exist scales which assess the arousal dimension (Eysenck and Eysenck, 1968), as well as related issues such as the degree of sensation seeking (Zuckerman et al., 1964). There are also scales which assess perceptual processing ability (Mehrabian, 1976; Sarason, 1972).

## 3.0 EXPERIMENTAL METHODOLOGY

The primary goal of this project was to develop a comprehensive CTS data base for use as standardization data. A secondary goal was to explore context variable and individual difference variable effects on CTS performance. The design of the study afforded the opportunity to collect data which would simultaneously address both of these goals.

Figure 3-1 presents the overall testing protocol for the primary and secondary testing efforts. The overall testing protocol consisted of two-hour testing sessions conducted once per day for ten days over a two-week testing cycle. Multiple workstations allowed collection of up to five (5) subjects' data per two-hour session. In the majority of cycles, four two-hour sessions were conducted each test day. Approximately 16 to 20 subjects were run during a two-week cycle. Numerous two-week testing cycles were conducted to collect the data necessary for this project. The Primary Study took place during the first nine days of each two-week cycle, and Secondary Study efforts were performed on the tenth day.

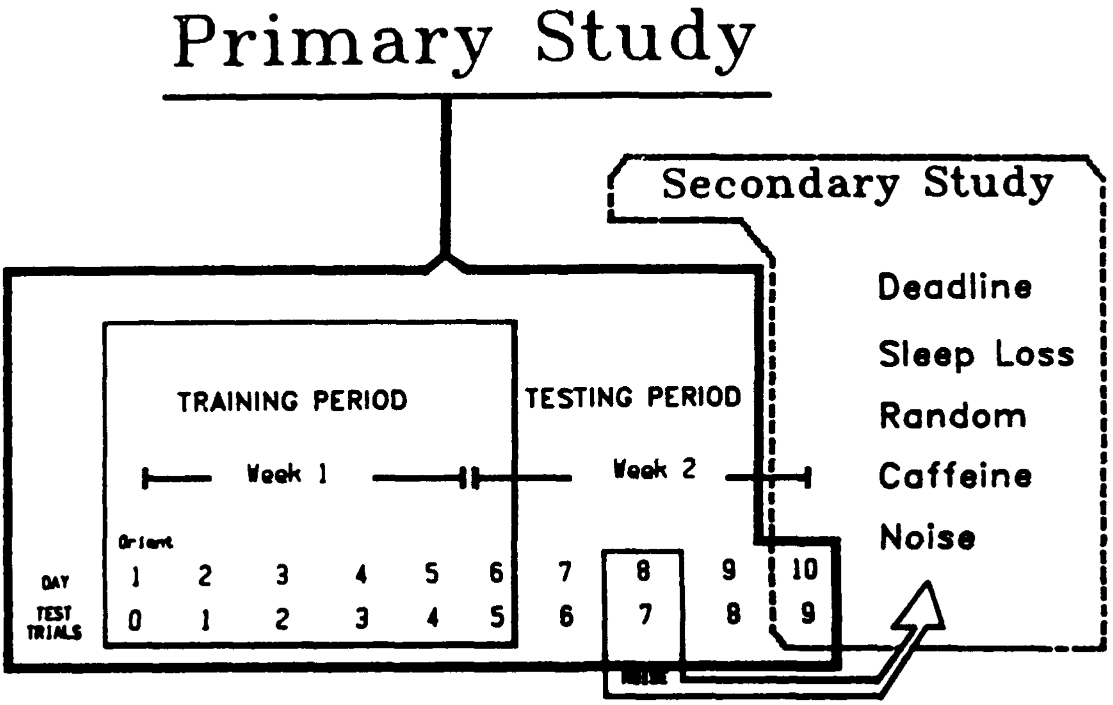


Figure 3-1. Testing Protocol for Primary and Secondary Studies.

### 3.1 Dependent Measures

The comprehensive data base includes performance, subjective assessment, and individual differences measures. The data is stored in a form that allows easy access by the Statistical Analysis System (SAS; SAS Institute, 1985).

### 3.1.1 Performance Measures

All tasks except Interval Production were run as standard three-minute trials under the subject-paced condition (CTS Menu Option 1), which places a 15-second limit on subject response time for the central processing tasks. The performance measures for the central processing tasks include the mean and standard deviation of response time for correct responses and counts of total stimuli, and correct and incorrect (both errors and missed) responses during each three-minute trial. These measures are also derived separately for those stimuli with positive ("YES") responses and negative ("NO") responses.

The performance measures for Unstable Tracking include the mean absolute error and total edge violations for the three-minute trial. Measures for Interval Production include the mean and standard deviation of the tapping intervals along with the variability score for the trial. See Shingledecker (1984) for additional details.

### 3.1.2 Subjective Workload Measure

Throughout both the Primary and Secondary Studies, subjects were asked to provide subjective assessments of the workload presented by the various CTS tasks. The Subjective Workload Assessment Technique (SWAT) was used to assess subjective workload. Based on conjoint measurement this technique constructs an interval scale for mental workload from ordinal rankings of events involving three hypothesized components of workload. These independent variables are an adaptation of the theoretical framework used by Sheridan and Simpson (1979) in developing a category scale for workload assessment. The dimensions used in SWAT are time load (T), mental effort load (E), and psychological stress load (S) (Reid, 1982; Reid et al., 1981).

The SWAT is a two-step process involving a scale development phase and an event scoring phase. During the scale development phase, subjects are asked to rank, from low to high, 27 cards representing all possible combinations of the three levels of Time, Effort and Stress. This is referred to as the "SWAT sort". Once a subject orders the dependent variable, an additive conjoint measurement composition rule for the ordered data is tested using various axiom tests. A scaling transformation is then computed to establish the interval scale for workload (Nygren, 1982).

The event scoring phase of the SWAT is an implementation of the scale as a dependent variable. During this phase, subjects rate the mental workload associated with a task by assigning a 1, 2 or 3 on each of the three dimensions. These values are defined by the same descriptors that were previously used for scale development. This rating is then converted to the scale value associated with this particular combination from the scale development phase. The event scoring should not interfere with the normal performance of the task, but should be made as temporally close to the events of interest as possible.

The unique aspect of SWAT is that it not only provides a means for obtaining an individual subject's workload ratings, but it also provides a method for establishing cross-subject comparability. Several research studies have been conducted to test the validity of the SWAT in both field and laboratory settings. In addition, work has been performed to evaluate different scaling approaches and SWAT administration methods.

### 3.1.3 Individual Difference Measures

Subjects in the Primary Study were also administered a battery of psychometric tests measuring individual difference dimensions that have a known or hypothesized relationship to performance or perceptual abilities. This battery included measures of generalized arousal (extraversion), stimulus screening, sensation seeking, test anxiety, clinical anxiety, and Type A behavior. Table 3-1 provides a legend identifying the variable names used.

**Generalized Arousal (Introversion-Extraversion).** The Eysenck Personality Inventory was used to assess generalized arousal (Eysenck and Eysenck, 1968). This dimension is believed to be directly related to brainstem reticular formation activity which is subsequently reflected in different levels of cortical arousal. Introverts are hypothesized to be higher in arousal than extraverts. This arousal difference often leads to one group or the other having a performance advantage depending on the environmental or task circumstances. Reviews of both the performance and psychophysiological literature generally support this theory. This inventory also provides the following subscales: **Neuroticism**, reflecting emotional responsivity; **Sociability**, reflecting level of interest in social affiliation; and **Impulsivity**, the propensity to respond quickly often without thought or reflection.

**Stimulus Screening.** Also related to the orienting reflex, as well as arousability, is the dimension of stimulus screening (Mehrabian, 1976). Stimulus screening refers to a biologically-based, perception-related dimension that reflects one's ability to screen relevant and irrelevant stimuli during information processing.

**Sensation Seeking.** Developed from early sensory deprivation and optimal level of arousal research, the sensation seeking scale (Zuckerman, 1979; Form V) assesses the degree to which people actively seek sensory stimuli to increase their stimulation level. This dimension has been related to orienting reflex diiferences (Zuckerman, 1972) and more recently to regulators of neurotransmitters (Murphy et al., 1977). This inventory also provides subscales assessing an individual's level of **Thrill and Adventure Seeking, Experience Seeking, Boredom Susceptibility**, and **Disinhibition**.

**Test Anxiety.** Test anxiety is a form of anxiety associated with demand for performance. One scale of test anxiety (Sarason, 1972) has shown negative correlations with performance efficiency, especially on vigilance and selective attention tasks.

**Impulsiveness.** Impulsiveness has been shown to be related to physiological processes, especially arousal mechanisms. Impulsiveness was measured with the Barratt Impulsiveness Scale (Barratt, 1965) which also provides subscales of **Motor, Cognitive** and **Non-Planning Impulsiveness**.

**Clinical Anxiety.** Clinical anxiety, in a more general sense, is simply termed **anxiety**, as opposed to more specialized forms such as test anxiety. Anxiety is known to disrupt motor performance and cognition. Anxiety is usually viewed as being either of a transient "state" form often due to situational factors or a more pervasive, protracted "trait" form. Both trait and state anxiety were assessed with the State-Trait Anxiety Inventory (Spielberger et al., 1970).

**Type A Behavior.** Type A Behavior refers to a specific coping style which has been linked to coronary prone behavior. This dimension is interesting for two reasons. First, it shows an apparent relationship to physiological processes, e.g., cardiovascular responses. Second, it appears to be related to highly organized, stressful, competitive, and overscheduled approaches to problem solving. This dimension was measured with the Jenkins Activity Survey (Jenkins et al., 1979).

**Intelligence.** While the theor^tical nature of general intelligence remains controversial, this dimension has been shown to be a mediating factor in the performance of many tasks. The **Wechsler Adult Intelligence Scale-Revised (WAIS-R)** was administered to all subjects in the study.

**Hardiness.** Some individuals appear to have greater resistance to the negative effects of life stress than others (Kobasa and Maddi, 1977). These individuals are hypothesized to score high on the Hardiness Scale and would be expected to have better general levels of adjustment and better levels of joh performance.

### 3.2 Facilities and Equipment

A three-room suite in the basement of Dale Hall at the University of Oklahoma was utilized for this study. One room served as a CTS data collection area, another room served as a data management/reduction area, and the third room was a psychophysiological testing area which served several ancillary testing purposes.

Installed in the data collection area were five workstations for subjects, each containing a color CRT monitor (Commodore Model 1702) and the three, standard response controllers designed for the CTS battery by the Workload and Ergonomics Branch at Wright-Patterson Air Force Base. These consisted of a tapping paddle controller box for the Interval Production task, a turn-pot controller box for the Unstable Tracking task, and a four-pushbutton controller box for the remaining central processing and input/perceptual tasks. Task learning aids were attached to the wall in front of each subject.

Installed immediately behind the subjects was the experimenter control station which included a test computer (Commodore 64) with two floppy disk drives (Commodore Model 1541) and a color CRT monitor (Commodore Model 1702) for each subject workstation. Epyx$_{TM}$ FastLoad$_{TM}$ cartridges were used to reduce disk access times during task loading and data storage.

The data management/reduction room was used for subject training and data reduction/transfer functions. Installed in this room was a terminal directly wired to the University IBM 3081 mainframe computer. This terminal provided direct access to larger computing capacity for data analysis and SWAT sorting analysis. Also contained in this room was a general preparation area and storage area for CTS testing supplies.

**Table 3-1. Description of Variable Names for Individual Differences Variables.**

| Variable Name | Description |
|---|---|
| INT-EXT | General Arousal |
| IE-SOC | Sociability Subscale |
| IE-IMP | Impulsivity Subscale |
| NEUROT | Neuroticism Subscale |
| STIMSCRE | Stimulus Screening |
| SENSSEEK | Sensation Seeking |
| SS-TAS | Thrill and Adventure Seeking Subscale |
| SS-EXPER | Experience Seeking Subscale |
| SS-BORED | Boredom Susceptibility Subscale |
| SS-DISIN | Disinhibition Subscale |
| TESTANX | Test Anxiety Scale |
| IMPULSE | Impulsiveness |
| IM-MOTOR | Motor Impulsivity Subscale |
| IM-COGN | Cognitive Impulsivity Subscale |
| IM-NPLAN | Non-Planning Impulsivity Subscale |
| STATE | State Anxiety |
| TRAIT | Trait Anxiety |
| JENKINS | Jenkins Activity Survey (Type A) |
| WAIS | Wechsler Adult Intelligence Scale-R |
| HARDY | Hardiness Scale |

The psychophysiological testing area served as an ancillary testing area for such activities as WAIS testing, interviewing, and applied testing applications.

The CTS Version 1.0 tasks were, in general, written in the BASIC programming language and then compiled. Additional software was developed during this project to automate the presentation sequence of the tasks and automatically label and store raw data in disk files. Software was also written to analyze and reduce the raw data, construct summary statistics files, and label and store these files. A variation of this software is now available in Version 2.0 of the CTS.

### 3.3 Primary Study Procedure

Subjects were generally scheduled in one of four testing session periods: 8:00-10:00 a.m., 10:00-12:00 a.m., 1:00-3:00 p.m., and 3:00-5:00 p.m. Four to five subjects were scheduled during each period for a total of sixteen to twenty subjects per day. On rare occasions, sessions were not filled and it was necessary to run an additional session between 5:00 and 7:00 p.m. Subjects attended a minimum of ten (10), two-hour sessions -- one per day, Monday through Friday, for two weeks. This two-week testing cycle was illustrated previously (Figure 3-1).

Subjects were seated in individual workstations facing CRT displays elevated to eye level. Controller boxes were placed on a table in front of the subjects. Subjects were instructed to use their right hand for responding with the controller boxes. For a few subjects an exception was made if the subject was left handed and felt that using the right hand would cause a noticeable decrement in performance. Also on the table were a pencil and SWAT rating recording materials. The workstations were separated by acoustic panels to reduce noise and subject interaction.

On Monday of the first week, each subject was oriented to the project, given an introduction to each of the CTS tasks, completed a SWAT Sort, and completed a battery of psychometric tests. Additional psychometric tests were administered in packets taken home and completed at leisure by the subjects. Approximately two hours of additional individual difference testing were scheduled and completed during the two-week period.

On the second through fifth days of the first week, subjects were given the first four training trials on the entire CTS battery. Monday of the second week was the fifth and last training trial. Stimulus presentation rates for the central processing tasks were subject dependent with a liberal response deadline of 15 seconds.

The sixth and eighth trials on Tuesday and Thursday of Week 2 were baseline experimental trials. Data on these days were collected under standard laboratory conditions, i.e., the same conditions imposed during training (including 15-second response deadlines). Data from trial seven (Wednesday of Week 2) for several subjects was collected under a noise stress condition. While subjects performed the CTS tasks on this day they were exposed to 75dB*A* (SPL - 0.0002 $d/cm^2$) background noise supplied by a tape recording of superimposed conversations and activity from two air traffic control centers.

The ninth trial involved CTS performance under various conditions: (1) sleep deprivation, (2) response deadlines, (3) caffeine exposure and (4) random task sequence (explained below). Results for these conditions including noise exposure are presented in Part II of this contract report.

A quasi-random sequence of the nine CTS tasks was constructed with the restrictions that no two highly difficult tasks (based on previous subjective evaluation; Shingledecker, 1984) were adjacent and that the input/output tasks were balanced within the sequence. The subsequent task order used for all of the test sessions was as follows:

**(1) Memory Search**
**(2) Interval Production**
**(3) Continuous Recall**
**(4) Linguistic Processing**
**(5) Probability Monitoring**
**(6) Grammatical Reasoning**
**(7) Mathematical Processing**
**(8) Unstable Tracking**
**(9) Spatial Processing.**

Once the CTS task sequence was determined, the workload levels of each task were presented in ascending order within each task. During each testing session, subjects were thus presented three-minute trials of each of the 25 CTS task-level combinations (three workload levels for eight tasks, plus the Interval Production task). To investigate task and level sequence effects, twelve subjects performed the various tasks in one alternative fixed order and nine other subjects performed the tasks in a second alternative fixed order. In addition, 45 subjects performed the CTS with tasks and levels completely randomized. In general, performance on all tasks except Unstable

Tracking was consistent with the established learning curve at that point (Trial 9). Performance on Unstable Tracking as measured by the number of Edge Violations improved substantially more than would be predicted by the learning curve. Details of this analysis are reported in Part II of this report.

Following each trial was a brief rest period during which data was stored on the diskette and the next task was prepared for presentation. These rest periods were approximately 1 to 1.5 minutes in length depending on the number of subject responses. Duing these rest periods each subject recorded a SWAT rating for the previous CTS task trial. A total of 25 SWAT ratings were thus recorded during each testing session. Total test session time ranged from one hour and forty-five minutes to two hours depending on the data storage time.

### 3.4 Subjects

Over the course of the study, data was collected on a total of 132 individuals. All subjects were recruited through posted announcements and were paid for their participation in the study. The overwhelming majority of subjects were undergraduate students attending the University of Oklahoma. All subjects reported 20/20 actual or corrected vision, no history of hearing impairment, and no current use of medication.

Eight of the subjects were non-U.S. citizens and were dropped from the final data analysis for the following reasons:

(1) performance of these subjects on some CTS tasks indicated poor vocabulary skills, a lack of understanding of the task or both,

(2) the SWAT sorts provided by these subjects possessed numerous axiom violations with the exception of a few subjects who provided iterative sorts, and

(3) the individual difference measures might not be appropriate for non-U.S. citizens.

Following removal of the eight non-U.S. subjects, the data set was screened for outliers. For the six central processing tasks (CR, GR, LP, MP, MS and SP), the overall mean response time for correct responses and the proportion correct were examined on Trials 6 and 8. In general, the distribution of mean response times was positively skewed and the distribution of proportion correct was negatively skewed (particularly for the easier tasks) as would be expected for these measures.

A listing was made of those subjects whose scores deviated from the mean by more than 4 standard deviations. From more than 8900 responses, a total of 49 such outliers were identified (all on the *poor performance* side of the mean). These are categorized by task and level in Table 3-2. A separate breakdown by subject and trial is given in Table 3-3.

Table 3-3 and an examination of the daily experiment log (indicating a lack of understanding of the tasks) provided sufficient evidence to remove subject #92 from all further analyses. The outlier data for the other subjects was very task and trial specific and no other justification existed for removal of any of these subjects from the total data base. However, some subjects were removed for specific task analyses as described in Section 4.

For the Interval Production task, the mean and standard deviation of the intervals was examined along with the variability scores. Only one subject (#124) consistently tapped at a rate less than one tap per second. Other subjects (#6 on Trial 8, #73 on Trial 8, #76 on Trial 6) performed quite poorly as indicated by their variability scores. These three subjects were retained in the overall data set but were removed for the summaries and analyses of the IP performance data.

**Table 3-2. Number of Outlier Points for Central Processing Tasks by Task and Level.**

| Measure | Mean RT | | | Proportion Correct | | | Total |
|---|---|---|---|---|---|---|---|
| Level | Low | Med | High | Low | Med | High | |
| Task | | | | | | | |
| CR | 1 | 1 | | 3 | | | 5 |
| GR | | | | 3 | 1 | | 4 |
| LP | 2 | | 3 | 3 | 1 | 1 | 10 |
| MP | 3 | 2 | 2 | | 2 | 1 | 10 |
| MS | 3 | 4 | 2 | 2 | 2 | 1 | 14 |
| SP | 2 | | | 2 | 1 | 1 | 6 |
| Total | 25 | | | 24 | | | 49 |

**Table 3-3. Number of Outliers for Central Processing Tasks by Subject and Trial.**

| Subject ID | Trial 6 | Trial 8 | Total |
|---|---|---|---|
| 2 | 1 (MP) | | 1 |
| 4 | 1 (SP) | | 1 |
| 5 | | 1 (SP) | 1 |
| 17 | 1 (SP) | 1 (MS) | 2 |
| 22 | 2 (LP,GR) | 1 (LP) | 3 |
| 23 | 3 (MS) | | 3 |
| 36 | 1 (SP) | 2 (MP) | 3 |
| 44 | | 2 (CR,MS) | 2 |
| 65 | 1 (MS) | 2 (MS) | 3 |
| 68 | | 1 (MS) | 1 |
| 76 | | 1 (LP) | 1 |
| 79 | 1 (LP) | | 1 |
| 92 | 12 (GR,LP,MP,MS) | 8 (GR,LP,MP,MS) | 20 |
| 122 | 1 (LP) | | 1 |
| 124 | 1 (CR) | 1 (SP) | 2 |
| 133 | | 1 (LP) | 1 |
| 138 | 1 (CR) | 2 (CR,SP) | 3 |
| **Total** | 26 | 23 | 49 |

Mean absolute error and total edge violations were examined for the Unstable Tracking task. No outliers were observed for mean absolute error. Subjects #5, #72, and #93 performed somewhat poorly with respect to edge violations and were removed for the summaries and analyses of the UT performance data.

From the outlier screening, only one male subject (#92) was dropped from all analyses due to poor performance (scores 4 to 9 standard deviations worse than the mean) on four of the nine tasks. Other subjects occasionally performed poorly (worse than 4 standard deviations from the mean) on individual tasks and trials but were included in the analyses and summaries due to their minimal impact on the overall results.

In summary, 123 subjects (28 women and 95 men) were included in the analyzed data sets. Male subjects ranged in age from 17 to 34 years ($\bar{x}$ = 22.0, $s$ = 4.1) and female subjects ranged from 18 to 43 ($\bar{x}$ = 22.8, $s$ = 6.6). Table 3-4 summarizes the psychometric test scores for the subjects. Detailed analyses of the psychometric data

## Table 3-4. Summary of Subject Information.

| Variable | Females | | Males | | Overall | |
|---|---|---|---|---|---|---|
| | $\bar{x}$ | $(s)$ | $\bar{x}$ | $(s)$ | $\bar{x}$ | $(s)$ |
| AGE | 22.8 | (6.6) | 22.0 | (4.0) | 22.2 | (4.7) |
| INT-EXT | 12.5 | (3.8) | 12.7 | (3.9) | 12.7 | (3.9) |
| IE-SOC | 7.4 | (2.5) | 6.9 | (2.7) | 7.0 | (2.6) |
| IE-IMP | 4.0 | (1.5) | 4.4 | (2.2) | 4.3 | (2.0) |
| NEUROT | 10.9 | (4.8) | 9.4 | (4.5) | 9.7 | (4.6) |
| STIMSCRE | -25.4 | (37.6) | 0.6 | (38.8) | -5.4 | (39.9) |
| SENSSEEK | 18.0 | (4.8) | 21.3 | (6.1) | 20.6 | (6.0) |
| SS-TAS | 6.3 | (2.1) | 7.6 | (1.9) | 7.3 | (2.0) |
| SS-EXPER | 5.8 | (2.2) | 5.4 | (2.1) | 5.5 | (2.1) |
| SS-BORED | 2.0 | (1.2) | 3.1 | (2.1) | 2.9 | (2.0) |
| SS-DISIN | 3.9 | (1.9) | 5.2 | (2.9) | 4.9 | (2.8) |
| TESTANX | 15.9 | (5.7) | 13.1 | (5.7) | 13.7 | (5.8) |
| IMPULSE | 55.0 | (11.8) | 55.5 | (14.1) | 55.4 | (13.6) |
| IM-MOTOR | 17.7 | (5.3) | 16.7 | (6.3) | 16.9 | (6.1) |
| IM-COGN | 17.1 | (4.3) | 17.8 | (5.5) | 17.6 | (5.3) |
| IM-NPLAN | 20.2 | (6.4) | 21.0 | (6.7) | 20.8 | (6.6) |
| STATE | 35.5 | (8.7) | 35.7 | (7.5) | 35.6 | (7.8) |
| TRAIT | 39.5 | (8.5) | 37.4 | (7.9) | 37.9 | (8.0) |
| JENKINS | 7.4 | (2.6) | 7.4 | (3.3) | 7.4 | (3.1) |
| WAIS | 111.5 | (14.5) | 113.5 | (12.0) | 113.1 | (12.5) |
| HARDY | | (.) | 0.0 | (2.8) | 0.0 | (2.8) |

(see Table 3-1 for legend of variable names)

and its relationship to the performance and SWAT data are provided in Part III of this report. Table 3-5 summarizes the number of female and male subjects in the Primary Study and the various Secondary studies.

Table 3-5. Number of Subjects for Each Test Condition.

| Condition | Females | Males | Total |
|---|---|---|---|
| Training/ Baseline | 28 | 95 | 123 |
| Sleep Loss | 21 | 20 | 41 |
| Deadline | 6 | 7 | 13 |
| Caffeine | 0 | 12 | 12 |
| Random | 0 | 66 | 66 |
| Noise | 28 | 50 | 78 |

### 3.5 Data Base Organization

All data is stored on the University of Oklahoma's IBM 3081 mainframe in individual SAS databases. There are three major divisions of data related to the training and baseline data: CTS performance data, SWAT data and subject/individual difference data. All performance data was reduced on the Commodore 64 using software developed by the Principal Investigators. Programs were developed to automatically sequence through the files on each data disk, compute the summary statistics and write them to a disk file in a format appropriate for SAS input. The summarized data was stored on Commodore diskettes and subsequently transferred to a VAX equivalent computer on the College of Engineering's Engineering Computer Network (ECN). From ECN, the data was transferred to the IBM 3081 through a Remote Job Entry link and used to generate the SAS databases. A summary of the databases, their contents and storage requirements is given in Table 3-6.

The variables and formats used in the various data bases are summarized in Table 3-7. All data bases contain the variables ID, GROUP, SUBJECT, GENDER, and PTYPE. In addition, the performance and SWAT data bases all contain TASK, LEVEL and TRIAL with the appropriate performance or SWAT data. Trials are numbered 1 through 10 with trials 1 through 5 for training, 6 and 8 for baseline, 7 for noise stress and 9 and 10 for secondary studies. For the central processing tasks, stimulus data was summarized for all correct responses (overall), for correct responses requiring a *YES*, *MATCH* or *SAME* response (positive, right keypad button) and for correct responses requiring a *NO, NON-MATCH* or *DIFFERENT* response (negative, left keypad button). For Interval Production, a second variability score was computed to examine a possible error in the CTS variability measure.

The subject data base SUBJDB.ALL is essentially a zero/one matrix of subject participation data identifying subject inclusion in training, baseline and the various stressor studies and inclusion in various SWAT solutions. The individual difference measures in SUBJDB are summarized in Table 3-1.

**Table 3-6. Summary of SAS Data Bases for Training and Baseline Data.**

| Name | Members | Description | Size (# Obs.) | |
|---|---|---|---|---|
| **CTS Performance and SWAT Data** | | | | |
| All Performance and SWAT databases have the following members:<br>ALL - Data for all 132 subjects<br>MASTER - Data for 123 subjects used in analyses | | | | |
| | | | **ALL** | **MASTER** |
| CRDB | | Continuous Recall Perf. Data | 3,687 | 3,426 |
| GRDB | | Grammatical Reasoning Perf. Data | 3,687 | 3,426 |
| IPDB | | Interval Production Perf. Data | 1,229 | 1,142 |
| LPDB | | Linguistic Processing Perf. Data | 3,687 | 3,426 |
| MPDB | | Mathematical Processing Perf. Data | 3,687 | 3,426 |
| MSDB | | Memory Search Perf. Data | 3,687 | 3,426 |
| PMDB | | Probability Monitoring Perf. Data | 3,687 | 3,426 |
| SPDB | | Spatial Processing Perf. Data | 3,687 | 3,426 |
| UTDB | | Unstable Tracking Perf. Data | 3,687 | 3,426 |
| SWATDB | | Subjective Workload Ratings | 28,025 | 25,875 |
| **Subject/Individual Difference Data** | | | | |
| SUBJDB | ALL | Subject Participation Data | 132 | |
| | PERSONAL | Personality Measures for All Subjects | 132 | |
| | IND_DIFF | Individual Differences Data for Subjects Included in Study | 123 | |

Table 3-7. Summary of Variable Names.

| Name | Description | Values | Format |
|---|---|---|---|
| **All Data Bases** | | | |
| ID | unique subject identifier | 1 to 150 | F3.0 |
| GROUP | subject group identifier | A to I | $1 |
| SUBJECT | subject initial code | - | $3 |
| GENDER | subject gender | F,M | $1 |
| PTYPE | subject prototype | T,E,S,X | $1 |
| **Performance and SWAT Data Bases** | | | |
| TASK | two-character task code | CR,GR, etc. | $2 |
| LEVEL | task difficulty level | 1,2,3 | $1 |
| TRIAL | trial number | 01 to 10 | $3 |
| **Performance Data - Central Processing Tasks** | | | |
| XXMNO* | RT mean, corr. resp., overall | | F5.0 |
| XXSDO | RT std. dev., corr. resp., overall | | F5.0 |
| XXPCO | prop. corr., overall (CORO/STIMO) | | F6.4 |
| XXSTIMO | number of stimuli, overall | | F3.0 |
| XXCORO | number correct, overall | | F3.0 |
| XXERRO | number errors, overall | | F3.0 |
| XXMNP | RT mean, corr. resp., pos. | | F5.0 |
| XXSDP | RT std. dev., corr. resp., pos. | | F5.0 |
| XXPCP | prop. corr., pos. (CORP/STIMP) | | F6.4 |
| XXSTIMP | number of stimuli, pos. | | F3.0 |
| XXCORP | number correct, pos. | | F3.0 |
| XXERRP | number errors, pos. | | F3.0 |
| XXMNN | RT mean, corr. resp., neg. | | F5.0 |
| XXSDN | RT std. dev., corr. resp., neg. | | F5.0 |
| XXPCN | prop. corr., neg. (CORN/STIMN) | | F6.4 |
| XXSTIMN | number of stimuli, neg. | | F3.0 |
| XXCORN | number correct, neg. | | F3.0 |
| XXERRN | number errors, neg. | | F3.0 |

* where XX is replaced by the task code, e.g., CRMNO

Table 3-7. Summary of Variable Names (continued).

| Name | Description | Values | Format |
|---|---|---|---|
| | **Interval Production Data Base** | | |
| IPINT | number of tapping intervals | | F4.0 |
| IPMN | tapping interval mean | | F4.0 |
| IPSD | tapping interval std. dev. | | F4.0 |
| IPVS1 | CTS variability score | | F5.1 |
| IPVS2 | alternative variability score | | F6.4 |
| | **Probability Monitoring Data Base** | | |
| PMCR | no. of correct signal detections | | F2.0 |
| PMMB | no. of missed signal biases | | F2.0 |
| PMFA | no. of false alarms | | F2.0 |
| PMRT1-3 | RT's for signals 1,2 and 3 | | F4.1 |
| PMRT | mean RT for detected signals | | F4.1 |
| PMPC | prop. correct signal detections | | F6.4 |
| | **Unstable Tracking Data Base** | | |
| UTMAE | mean absolute error | | F5.1 |
| UTEV | number of edge violations | | F4.0 |
| | **SWAT Data Base** | | |
| TIME | rating on Time | 1,2,3 | $1 |
| EFFORT | rating on Effort | 1,2,3 | $1 |
| STRESS | rating on Stress | 1,2,3 | $1 |
| SWAT | scaled SWAT rating | 0 to 100 | F5.1 |
| | **Subject Data Base** | | |
| PSOLN | incl. in prototype solution | 0,1 | $1 |
| WGSOLN | incl. in whole group solution | 0,1 | $1 |
| BASE | incl. in data analyses | 0,1 | $1 |
| NOISE | incl. in noise study | 0,1 | $1 |
| SLEEP | incl. in sleep loss study | 0,1 | $1 |
| DEADLINE | incl. in deadline study | 0,1 | $1 |
| CAFFEINE | incl. in caffeine study | 0,1 | $1 |
| RANDOM | incl. in random seq. study | 0,1 | $1 |

## 4.0 PERFORMANCE DATA

The performance results will be presented separately for each task followed by an analysis of intertask relationships. For the central processing tasks, each trial involved the presentation of 20 to 200 individual stimuli. For these tasks, response time and accuracy were the primary measures recorded for each stimulus. The proportion of correct responses was computed by dividing the number correct by the total number of stimuli for the trial. The total number of incorrect responses was derived by subtracting the number correct from the total number of stimuli. The number of missed responses equals the number of incorrect responses minus the number of errors. For the non-central processing tasks, appropriate measures of speed and/or accuracy were used. A summary of the measures computed by the CTS Version 1.0 "STATISTICS" option was provided in Table 3-7.

For each task, a separate factor analysis was performed which included all relevant measures for that task. In general, for the discrete response central processing tasks, two factors were isolated, one related to speed of response (mean and standard deviation of response times, number of stimuli presented) and the other related to accuracy of response (proportion correct, number correct, number of errors). For this reason, the overall mean response time for correct responses (XXMNO) and the overall proportion correct (XXPCO) for the central processing tasks were selected for graphical presentation and analysis in this report. Other relevant variables are included in the tabular summaries. For the other tasks, all relevant variables are presented and analyzed.

### 4.1 Continuous Recall Task

The means and standard deviations for the Continuous Recall performance measures are presented in Table 4-1 for Trials 6 and 8. Overall mean response time (RT) and percentage correct (PC) are presented in Figures 4-1 and 4-2 with the intertrial correlations represented by the *r* values in the figures. Univariate summaries of the overall response time and proportion correct measures for the three difficulty levels are provided in Appendix A-1. Data for these summaries included the Trial 6 and Trial 8 data for all 123 subjects and thus provided pooled estimates of the various parameters.

Table 4-1. Means (Standard Deviations) of Performance Measures, Continuous Recall - Baseline Trials.

| Level | Low | | | Medium | | | High | | |
|---|---|---|---|---|---|---|---|---|---|
| **Trial** | **06** | **08** | **Both** | **06** | **08** | **Both** | **06** | **08** | **Both** |
| MNO | 1010 | 911 | 960 | 2126 | 2084 | 2105 | 3184 | 2971 | 3077 |
| | (341) | (287) | (318) | (788) | (870) | (829) | (1919) | (1849) | (1884) |
| SDO | 508 | 445 | 476 | 928 | 910 | 919 | 1246 | 1174 | 1210 |
| | (297) | (250) | (276) | (481) | (500) | (490) | (705) | (630) | (668) |
| PCO | 96 | 96 | 96 | 86 | 88 | 87 | 73 | 73 | 73 |
| | (5) | (5) | (5) | (14) | (13) | (13) | (13) | (13) | (13) |
| STIMO | 147 | 159 | 153 | 81 | 84 | 83 | 65 | 70 | 67 |
| | (34) | (35) | (35) | (36) | (35) | (36) | (42) | (43) | (43) |
| MNP | 875 | 800 | 838 | 1973 | 1904 | 1938 | 3061 | 2871 | 2966 |
| | (291) | (266) | (281) | (762) | (770) | (765) | (1961) | (1881) | (1920) |
| SDP | 367 | 332 | 350 | 814 | 803 | 808 | 1086 | 1079 | 1083 |
| | (225) | (226) | (226) | (462) | (451) | (455) | (749) | (693) | (720) |
| PCP | 95 | 95 | 95 | 81 | 83 | 82 | 61 | 61 | 61 |
| | (6) | (7) | (6) | (19) | (18) | (19) | (22) | (21) | (22) |
| MNN | 1136 | 1016 | 1076 | 2300 | 2275 | 2288 | 3286 | 3052 | 3169 |
| | (402) | (329) | (371) | (874) | (1060) | (970) | (1989) | (1873) | (1931) |
| SDN | 564 | 488 | 526 | 962 | 922 | 942 | 1254 | 1168 | 1211 |
| | (350) | (283) | (320) | (548) | (554) | (550) | (721) | (652) | (687) |
| PCN | 97 | 97 | 97 | 92 | 94 | 93 | 84 | 83 | 84 |
| | (6) | (6) | (6) | (13) | (11) | (12) | (14) | (16) | (15) |

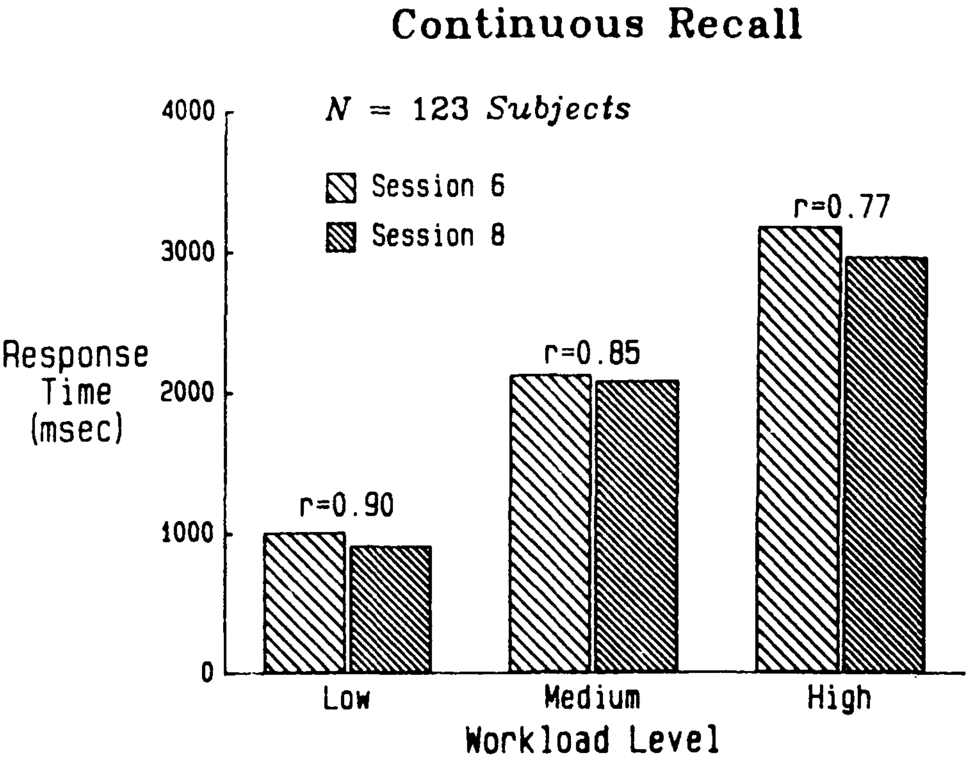


Figure 4-1. Mean Response Time for Continuous Recall - Trials 6 and 8.

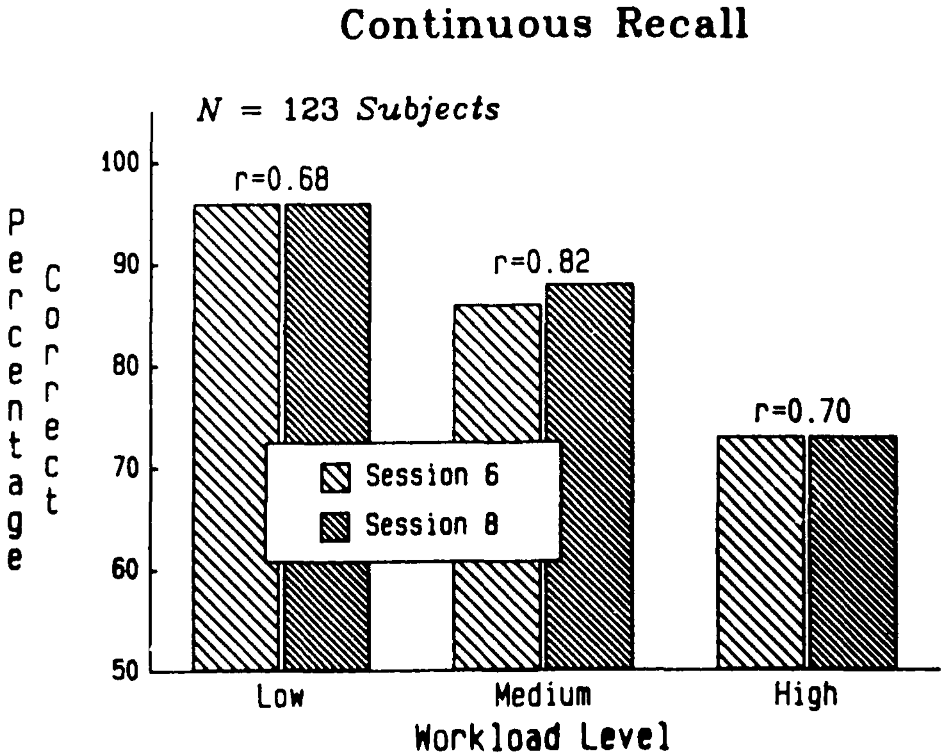


Figure 4-2. Percentage Correct for Continuous Recall - Trials 6 and 8.

Response times were approximately 1, 2 and 3 seconds for the low, medium and high difficulty levels with standard deviations that increased substantially with increasing difficulty. Proportion correct ranged from 0.96 to 0.73 with the same increase in standard deviation with increasing difficulty. For all three levels, times were faster but proportion correct was lower for the *MATCH* (right button) response› compared with the *NON-MATCH* (left button) responses. These differences were not analyzed for statistical significance.

The $r$ values for response time were large (0.77 to 0.90), indicating high stability in performance across subjects. While not as large, the $r$ values for percentage correct (0.68 to 0.82) indicated moderate stability across subjects. In general, the intertrial correlations for percentage correct were substantially lower than those for response time on all central processing tasks due to the ceiling effect (maximum of 100% correct) and the fact that most subjects performed at a high accuracy level.

### 4.1.1 Level and Trial Analyses

Analysis of variance was performed to verify the difficulty level manipulation and examine any trial differences using the following model:

$$RT_{ijk}, PC_{ijk} = \mu + L_i + T_j + LT_{ij} + S_k + LS_{ik} + TS_{jk} + \varepsilon_{ijk}$$

where:

$RT_{ijk}$ = Response Time

$PC_{ijk}$ = Proportion Correct

$L_i$ = Level, $i = 1,2,3$

$T_j$ = Trial (6 vs. 8), $j = 1,2$

$S_k$ = Subject, $k = 1, ..., 123$

$\varepsilon_{ijk}$ = Error (*LTS* interaction + random error; Winer, 1971, p. 378).

The results of the analyses for response time and proportion correct are summarized in Table 4-2. The $R^2$ values were 0.96 for both the RT and PC models. A Tukey studentized range test using $\alpha = 0.01$ demonstrated that all three difficulty levels differed significantly for both RT and PC. The mean response time for Trial 8 was significantly lower ($p = 0.008$) than the RT for Trial 6 with the largest difference occurring at the high difficulty level. This indicated continued improvement in RT even after substantial training on this task. There was no overall difference between trials for proportion correct, but a marginally significant Level by Trial interaction existed due to the slightly improved performance on Trial 8 at the medium level.

Table 4-2. ANOVA Summary for Level and Trial Effects, Continuous Recall - Baseline Trials.

| Response Time | | | | | |
|---|---|---|---|---|---|
| Source | DF | Sum of Squares (x1000) | Mean Square (x1000) | F | p > F |
| Level (L) | 2 | 552521 | 276260 | 147.74 | 0.0001 |
| Trial (T) | 1 | 2566 | 2566 | 7.27 | 0.0080 |
| L by T | 2 | 934 | 467 | 1.58 | 0.2085 |
| Subject (S) | 122 | 487207 | 3993 | 13.49 | 0.0001 |
| L by S | 244 | 456249 | 1870 | 6.32 | 0.0001 |
| T by S | 122 | 43062 | 353 | 1.19 | 0.1255 |
| Error | 244 | 72235 | 296 | | |

| Proportion Correct | | | | | |
|---|---|---|---|---|---|
| Source | DF | Sum of Squares (x0.001) | Mean Square (x0.001) | F | p > F |
| Level (L) | 2 | 6954 | 3477 | 278.43 | 0.0001 |
| Trial (T) | 1 | 6 | 6 | 1.73 | 0.1913 |
| L by T | 2 | 14 | 7 | 2.61 | 0.0754 |
| Subject (S) | 122 | 4900 | 40 | 14.58 | 0.0001 |
| L by S | 244 | 3047 | 12 | 4.53 | 0.0001 |
| T by S | 122 | 450 | 4 | 1.34 | 0.0286 |
| Error | 244 | 672 | 3 | | |

### 4.1.2 Gender and Prototype Analyses

To identify performance differences related to gender or prototype, additional analyses were performed by further partitioning of the subject variability. The previously mentioned ANOVA model (Section 4.1.1) was used with the addition of the factor GENDER (or PROTOTYPE) and its interactions with Level and with Trial. The Subject factor was thus nested within the grouping variable (gender or prototype).

The performance measures are presented separately for men and women in Table 4-3 and Figures 4-3 and 4-4. Men responded significantly faster (p = 0.009) than women, particularly at the high level (2794 vs. 4040 msec) with a significant (p < 0.0001) Level by Gender interaction. Although women were slightly more accurate (2% to 5%) at all levels, the difference was marginally significant (p = 0.059) with no

## Table 4-3. Means (Standard Deviations) of Performance Measures by Gender, Continuous Recall - Baseline Trials.

| Level | Low | | | Medium | | | High | | |
|---|---|---|---|---|---|---|---|---|---|
| Gender | Fem | Male | Both | Fem | Male | Both | Fem | Male | Both |
| MNO | 958<br>(353) | 961<br>(308) | 960<br>(318) | 2199<br>(825) | 2078<br>(830) | 2105<br>(829) | 4040<br>(2299) | 2794<br>(1645) | 3077<br>(1884) |
| SDO | 468<br>(255) | 479<br>(282) | 476<br>(276) | 934<br>(376) | 915<br>(519) | 919<br>(490) | 1469<br>(672) | 1133<br>(649) | 1210<br>(668) |
| PCO | 98<br>(2) | 96<br>(5) | 96<br>(5) | 89<br>(12) | 86<br>(14) | 87<br>(13) | 77<br>(12) | 72<br>(13) | 73<br>(13) |
| STIMO | 154<br>(38) | 153<br>(34) | 153<br>(35) | 77<br>(28) | 84<br>(37) | 83<br>(36) | 51<br>(33) | 72<br>(44) | 67<br>(43) |
| MNP | 827<br>(286) | 841<br>(280) | 838<br>(281) | 2024<br>(811) | 1913<br>(752) | 1938<br>(765) | 3923<br>(2443) | 2684<br>(1640) | 2966<br>(1920) |
| SDP | 325<br>(191) | 357<br>(235) | 350<br>(226) | 816<br>(433) | 806<br>(463) | 808<br>(455) | 1245<br>(730) | 1035<br>(712) | 1083<br>(720) |
| PCP | 98<br>(3) | 95<br>(7) | 95<br>(6) | 85<br>(18) | 81<br>(19) | 82<br>(19) | 66<br>(21) | 60<br>(21) | 61<br>(22) |
| MNN | 1079<br>(418) | 1075<br>(358) | 1076<br>(371) | 2384<br>(885) | 2259<br>(994) | 2288<br>(970) | 4156<br>(2367) | 2878<br>(1682) | 3169<br>(1931) |
| SDN | 522<br>(307) | 527<br>(324) | 526<br>(320) | 960<br>(413) | 937<br>(585) | 942<br>(550) | 1458<br>(640) | 1138<br>(685) | 1211<br>(687) |
| PCN | 98<br>(2) | 97<br>(7) | 97<br>(6) | 94<br>(9) | 93<br>(13) | 93<br>(12) | 87<br>(13) | 83<br>(16) | 84<br>(15) |

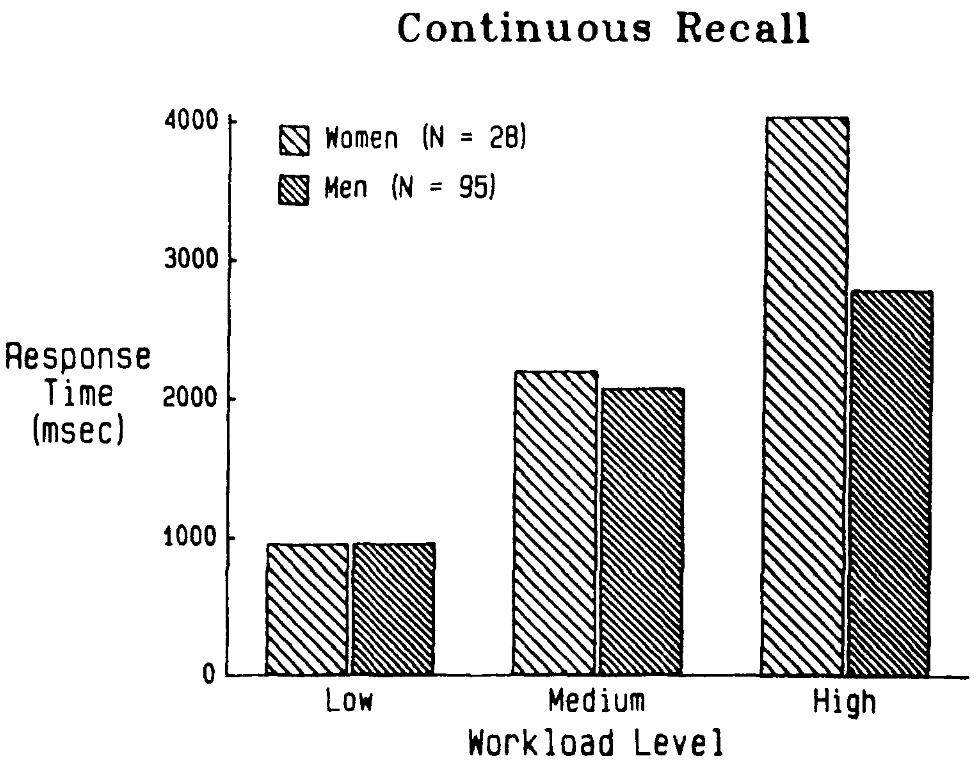


Figure 4-3. Mean Response Time for Continuous Recall - Men vs. Women.

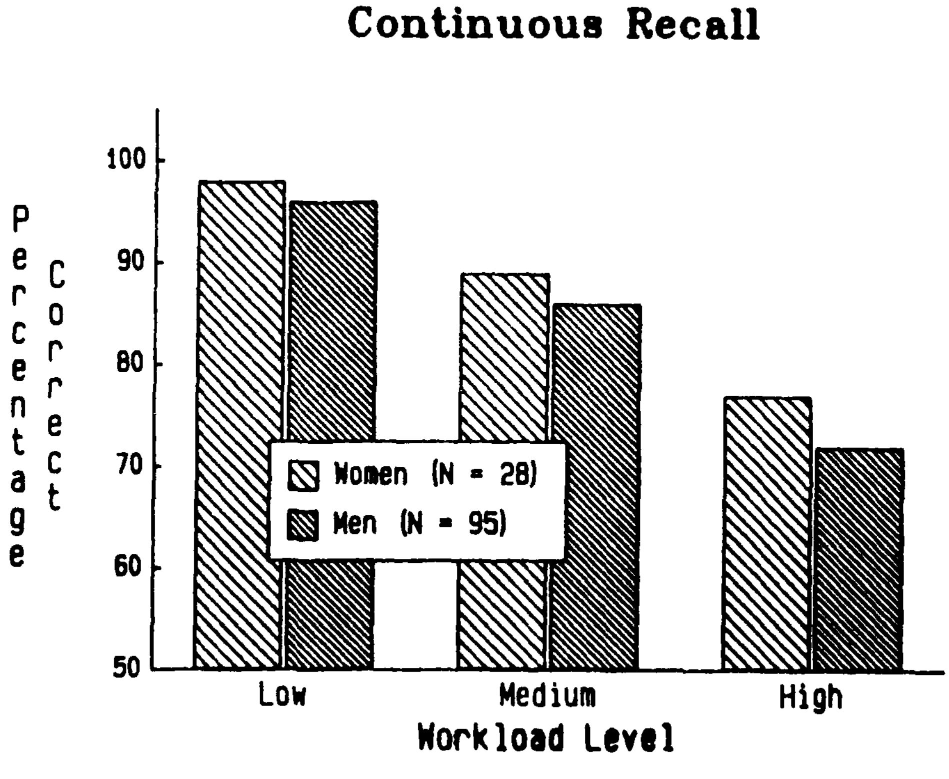


Figure 4-4. Percentage Correct for Continuous Recall - Men vs. Women.

Level by Gender interaction. There was no significant Trial by Gender interaction for either response variable.

Although the difference was not significant, the Effort prototype group tended to have faster response times than the other groups. In addition, there was a significant ($p = 0.01$) Trial by Prototype interaction for RT. There were no differences for the proportion correct measure. Refer to Section 5.2 for a discussion of the prototype grouping.

### 4.1.3 Training Data

The means and standard deviations of the major performance measures for Continuous Recall for training Trials 1 through 5 are presented by difficulty level in Table 4-4. Response time and percentage correct are plotted in Figures 4-5 and 4-6. There was significant improvement in both speed and accuracy during training with the largest improvement from Trial 1 to Trial 2. As shown in the figures, asymptotic performance was achieved on different trials depending on the difficulty level and the particular performance measure.

Analysis of variance was used to determine significance between trials for RT and PC using the model presented in Section 4.1.1. A summary of the ANOVA results is presented in Table 4-5. Due to the significant Trial by Level interactions ($p = 0.005$ for RT, $p < 0.0001$ for PC), separate analyses were performed for each level using a reduced model involving only the trial and subject effects. The results of Tukey studentized range tests are summarized in Table 4-6. In all cases, there were no significant differences between Trials 4 and 5 although the improvement trend continued.

## Table 4-4. Means (Standard Deviations) of Performance Measures, Continuous Recall - Training Trials.

| Var. | Level | Trial | | | | |
|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 |
| MNO | Low | 1550 ( 403) | 1308 ( 394) | 1214 ( 334) | 1091 ( 324) | 1055 ( 343) |
| | Med | 2715 (1034) | 2392 ( 898) | 2359 ( 901) | 2229 ( 904) | 2187 ( 809) |
| | High | 3328 (1797) | 2939 (1466) | 3056 (1777) | 2998 (1837) | 3100 (1874) |
| SDO | Low | 860 ( 378) | 634 ( 276) | 582 ( 257) | 509 ( 236) | 512 ( 282) |
| | Med | 1516 ( 647) | 1096 ( 474) | 1083 ( 520) | 933 ( 384) | 942 ( 436) |
| | High | 1561 ( 777) | 1308 ( 642) | 1255 ( 664) | 1222 ( 691) | 1250 ( 689) |
| PCO | Low | 91 ( 9) | 94 ( 7) | 96 ( 6) | 96 ( 7) | 96 ( 6) |
| | Med | 74 (15) | 79 (15) | 83 (14) | 84 (15) | 87 (13) |
| | High | 65 (10) | 67 ( 9) | 69 (12) | 70 (13) | 72 (12) |
| STIMO | Low | 96 (22) | 115 (26) | 124 (27) | 137 (32) | 141 (32) |
| | Med | 61 (25) | 70 (28) | 73 (31) | 79 (37) | 78 (33) |
| | High | 55 (28) | 61 (31) | 64 (36) | 68 (43) | 66 (40) |
| MNP | Low | 1282 ( 339) | 1104 ( 336) | 1042 ( 291) | 954 ( 296) | 911 ( 300) |
| | Med | 2295 ( 997) | 2123 ( 868) | 2138 ( 882) | 2044 ( 895) | 1998 ( 817) |
| | High | 3063 (1853) | 2747 (1576) | 2851 (1785) | 2814 (1780) | 2953 (1915) |
| MNN | Low | 1788 ( 497) | 1491 ( 450) | 1371 ( 386) | 1218 ( 369) | 1194 ( 408) |
| | Med | 3086 (1194) | 2660 (1019) | 2562 ( 972) | 2414 ( 954) | 2351 ( 833) |
| | High | 3476 (1818) | 3078 (1463) | 3183 (1816) | 3122 (1935) | 3199 (1915) |

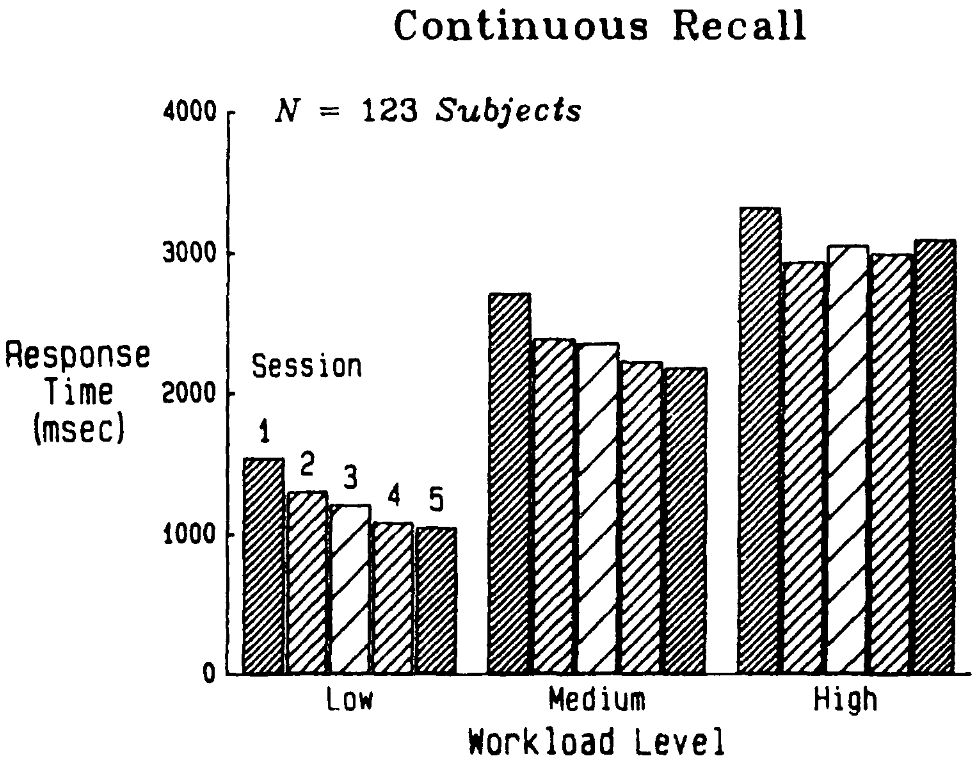


Figure 4-5. Mean Response Time for Continuous Recall - Trials 1 through 5.

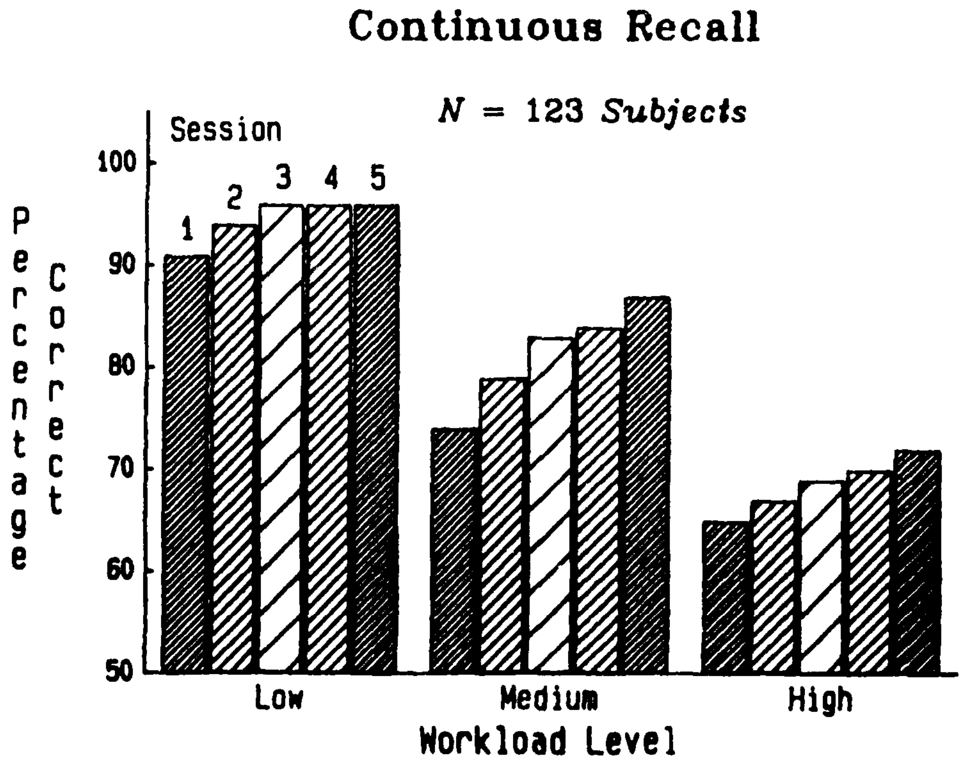


Figure 4-6. Percentage Correct for Continuous Recall - Trials 1 through 5.

### Table 4-5. ANOVA Summary for Level and Trial Effects, Continuous Recall - Training Trials.

| Var. | Model | Level | | Trial | | Level * Trial | |
|---|---|---|---|---|---|---|---|
| | $R^2$ | $F_{(2,244)}$ | $p>F$ | $F_{(4,488)}$ | $p>F$ | $F_{(8,976)}$ | $p>F$ |
| CRMNO | 0.91 | 163.26 | * | 18.73 | * | 2.75 | .0052 |
| CRPCO | 0.91 | 508.30 | * | 69.06 | * | 7.90 | * |

* p < 0.0001

### Table 4-6. Significant (α = .01) Trial Differences by Level.

| Var. | Level | $F_{(4,488)}$ | Trial |
|---|---|---|---|
| MNO | L | 128.33 | $1\ \ 2\ \ 3\ \ \underline{4\ \ 5}$ |
| | M | 18.27 | $1\ \ \underline{2\ \ 3\ \ 4\ \ 5}$ |
| | H | 2.95 | $\underline{1\ \ 5\ \ 3\ \ 4\ \ 2}$ |
| PCO | L | 24.17 | $1\ \ \underline{2\ \ 4\ \ 3\ \ 5}$ |
| | M | 54.03 | $1\ \ 2\ \ \underline{3\ \ \underline{4}}\underline{\ \ 5}$ |
| | H | 16.88 | $\underline{1\ \ \underline{2}}\ \ \underline{\underline{3}\ \ \underline{4}}\ \ \underline{5}$ |

## 4.2 Grammatical Reasoning Task

The means and standard deviations for the Grammatical Reasoning performance measures are presented in Table 4-7 for Trials 6 and 8. Overall mean response time (RT) and *percentage correct* (PC) *are presented in* Figures 4-7 and 4-8 with the inter-trial correlations represented by the $r$ values in the figures. Univariate summaries of the overall response time and proportion correct measures for the three difficulty levels are provided in Appendix A-2. Data for these summaries included the Trial 6 and Trial 8 data for all 123 subjects.

Response times for GR were the longest of all the central processing tasks at 3.3, 5.6 and 7.5 seconds for the low, medium and high difficulty levels respectively. The standard deviation of response time increased substantially with increasing difficulty. Proportion correct ranged from 0.93 to 0.85 with the same increase in standard deviation with increasing difficulty. As with CR, response times were faster but proportion correct was lower for the *MATCH* (right button) responses compared with the *NON-MATCH* (left button) responses.

The $r$ values for response time were again fairly large (0.75 to 0.82), indicating good stability in performance across subjects. The $r$ values for percentage correct (0.63 to 0.79) indicated moderate stability across subjects for the Grammatical Reasoning task.

### 4.2.1 Level and Trial Analyses

Analysis of variance was performed to verify the difficulty level manipulation and examine any trial differences using the model presented in Section 4.1.1. The results of the analyses for response time and proportion correct are summarized in Table 4-8. The model $R^2$ values were 0.98 for RT and 0.91 for PC. All three difficulty levels differed significantly for RT. With respect to PC, the low (0.93) and medium (0.91) levels did not differ but both differed from the high level (0.85). The mean response time for Trial 8 was significantly lower (p = 0.006) than the RT for Trial 6 with larger differences as the difficulty level increased. This indicated continued improvement on GR beyond the five training trials. There was no overall difference between trials for proportion correct and no significant Level by Trial interaction.

## Table 4-7. Means (Standard Deviations) of Performance Measures, Grammatical Reasoning - Baseline Trials.

| Level | Low | | | Medium | | | High | | |
|---|---|---|---|---|---|---|---|---|---|
| Trial | 06 | 08 | Both | 06 | 08 | Both | 06 | 08 | Both |
| MNO | 3289<br>(1159) | 3215<br>(1117) | 3252<br>(1136) | 5755<br>(1527) | 5502<br>(1497) | 5628<br>(1514) | 7654<br>(1776) | 7291<br>(1844) | 7472<br>(1816) |
| SDO | 1515<br>(687) | 1488<br>(691) | 1501<br>(688) | 1749<br>(608) | 1707<br>(637) | 1728<br>(622) | 2064<br>(701) | 2100<br>(731) | 2082<br>(715) |
| PCO | 93<br>(9) | 93<br>(9) | 93<br>(9) | 91<br>(11) | 91<br>(12) | 91<br>(11) | 85<br>(15) | 86<br>(14) | 85<br>(15) |
| STIMO | 56<br>(17) | 57<br>(18) | 56<br>(17) | 32<br>(11) | 33<br>(12) | 32<br>(12) | 23<br>(9) | 25<br>(11) | 24<br>(10) |
| MNP | 3212<br>(1127) | 3167<br>(1189) | 3190<br>(1156) | 5647<br>(1573) | 5378<br>(1600) | 5513<br>(1589) | 7531<br>(1806) | 7<br>(1792) | 7286<br>(1812) |
| SDP | 1512<br>(746) | 1432<br>(709) | 1472<br>(727) | 1704<br>(658) | 1629<br>(734) | 1666<br>(697) | 2063<br>(948) | 2059<br>(964) | 2061<br>(954) |
| PCP | 92<br>(11) | 93<br>(9) | 92<br>(10) | 90<br>(15) | 90<br>(14) | 90<br>(15) | 84<br>(17) | 85<br>(17) | 85<br>(17) |
| MNN | 3353<br>(1227) | 3273<br>(1180) | 3313<br>(1202) | 5842<br>(1598) | 5644<br>(1525) | 5743<br>(1562) | 7779<br>(1984) | 7537<br>(2051) | 7658<br>(2017) |
| SDN | 1476<br>(723) | 1492<br>(751) | 1484<br>(736) | 1711<br>(704) | 1695<br>(719) | 1703<br>(710) | 1896<br>(761) | 1935<br>(722) | 1916<br>(740) |
| PCN | 94<br>(8) | 93<br>(11) | 94<br>(9) | 93<br>(11) | 92<br>(12) | 93<br>(11) | 86<br>(18) | 87<br>(15) | 86<br>(16) |

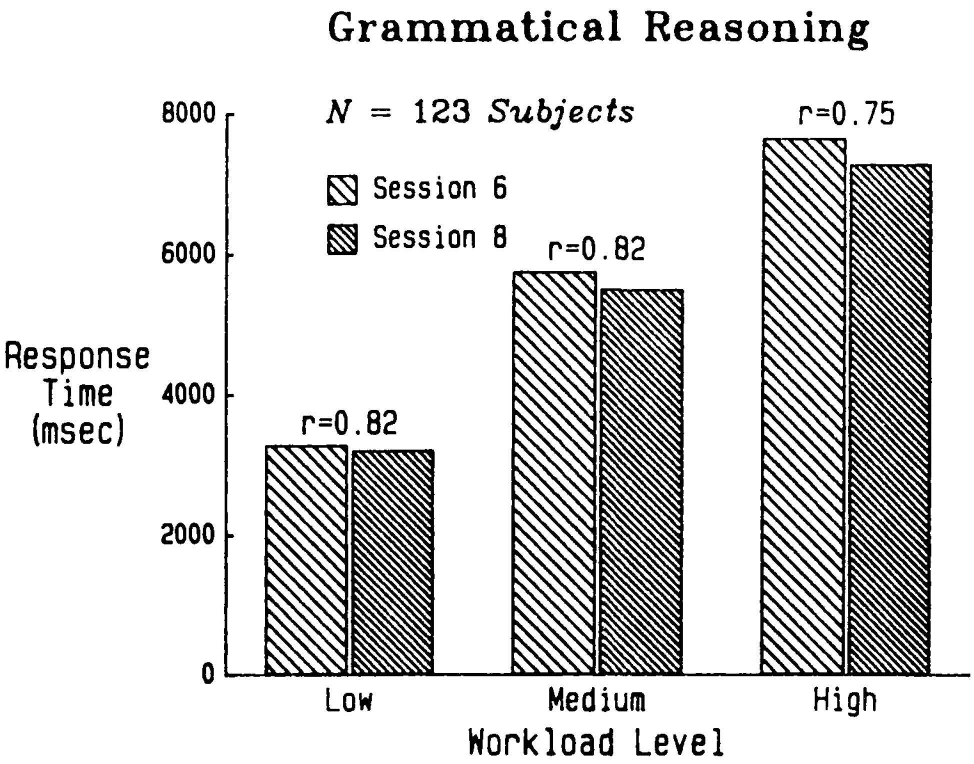


**Figure 4-7. Mean Response Time for Grammatical Reasoning - Trials 6 and 8.**

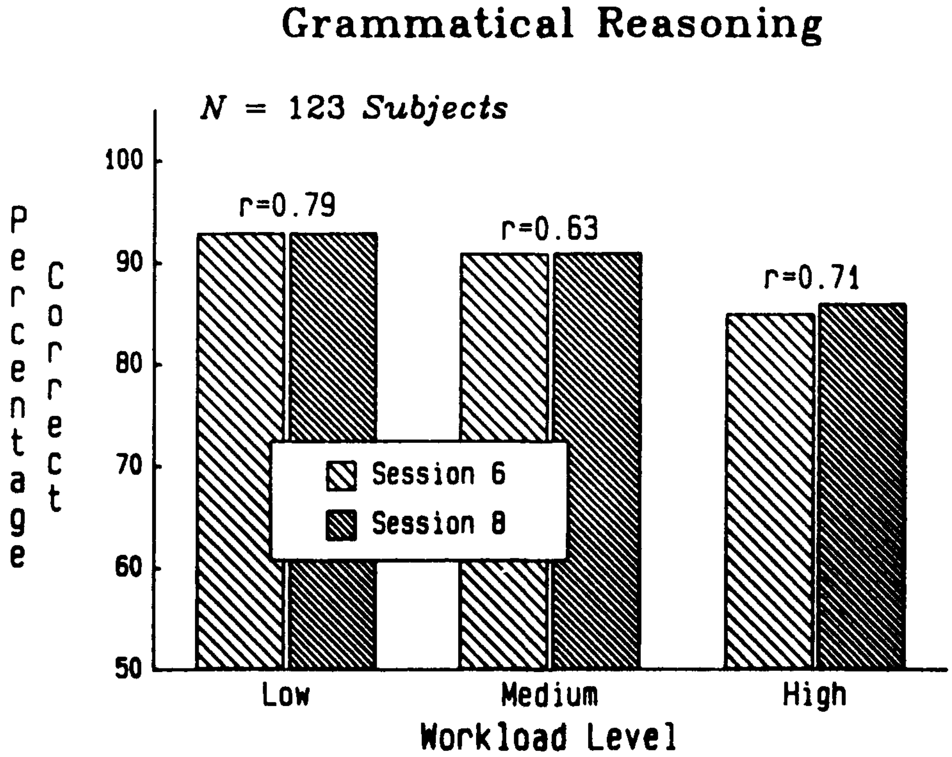


**Figure 4-8. Percentage Correct for Grammatical Reasoning - Trials 6 and 8.**

Table 4-8. ANOVA Summary for Level and Trial Effects, Grammatical Reasoning - Baseline Trials.

| Response Time | | | | | |
|---|---|---|---|---|---|
| Source | DF | Sum of Squares (x1000) | Mean Square (x1000) | F | p > F |
| Level (L) | 2 | 2202369 | 1101184 | 1165.88 | 0.0001 |
| Trial (T) | 1 | 9724 | 9724 | 12.49 | 0.0006 |
| L by T | 2 | 2629 | 1315 | 3.96 | 0.0202 |
| Subject (S) | 122 | 1267295 | 10388 | 31.31 | 0.0001 |
| L by S | 244 | 230460 | 945 | 2.85 | 0.0001 |
| T by S | 122 | 94998 | 779 | 2.35 | 0.0001 |
| Error | 244 | 80956 | 332 | | |

| Proportion Correct | | | | | |
|---|---|---|---|---|---|
| Source | DF | Sum of Squares (x0.001) | Mean Square (x0.001) | F | p > F |
| Level (L) | 2 | 817 | 409 | 44.35 | 0.0001 |
| Trial (T) | 1 | 0 | 0 | 0.08 | 0.7728 |
| L by T | 2 | 7 | 3 | 0.87 | 0.4214 |
| Subject (S) | 122 | 6559 | 54 | 13.41 | 0.0001 |
| L by S | 244 | 2247 | 9 | 2.30 | 0.0001 |
| T by S | 122 | 596 | 5 | 1.22 | 0.0974 |
| Error | 244 | 978 | 4 | | |

### 4.2.2 Gender and Prototype Analyses

As described in Section 4.1.2 for CR, further analyses of gender and prototype differences were performed. The performance measures are presented separately for men and women in Table 4-9 and Figures 4-9 and 4-10. There were no differences in response time or proportion correct between men and women and no significant interactions involving gender.

As with CR, the Effort prototype group tended to have faster response times than the other groups but the difference was not significant. There was a marginally significant (p=0.06) Trial by Prototype interaction for RT. There were no differences for the proportion correct measure. Refer to Section 5.2 for a discussion of the prototype grouping.

## Table 4-9. Means (Standard Deviations) of Performance Measures by Gender, Grammatical Reasoning - Baseline Trials.

| Level | Low | | | Medium | | | High | | |
|---|---|---|---|---|---|---|---|---|---|
| Gender | Fem | Male | Both | Fem | Male | Both | Fem | Male | Both |
| MNO | 3194<br>(1244) | 3269<br>(1106) | 3252<br>(1136) | 5442<br>(1052) | 5683<br>(1624) | 5628<br>(1514) | 7153<br>(1252) | 7567<br>(1944) | 7472<br>(1816) |
| SDO | 1404<br>(689) | 1530<br>(686) | 1501<br>(688) | 1590<br>(531) | 1769<br>(642) | 1728<br>(622) | 1931<br>(541) | 2127<br>(754) | 2082<br>(715) |
| PCO | 92<br>(13) | 93<br>(7) | 93<br>(9) | 92<br>(13) | 91<br>(11) | 91<br>(11) | 88<br>(13) | 84<br>(15) | 85<br>(15) |
| STIMO | 58<br>(17) | 56<br>(17) | 56<br>(17) | 32<br>(6) | 33<br>(13) | 32<br>(12) | 24<br>(4 ) | 24<br>(12) | 24<br>(10) |
| MNP | 3162<br>(1314) | 3198<br>(1109) | 3190<br>(1156) | 5330<br>(1141) | 5566<br>(1698) | 5513<br>(1589) | 7036<br>(1317) | 7359<br>(1931) | 7286<br>(1812) |
| SDP | 1376<br>(692) | 1500<br>(737) | 1472<br>(727) | 1613<br>(628) | 1682<br>(716) | 1666<br>(697) | 1888<br>(662) | 2113<br>(1021) | 2061<br>(954) |
| PCP | 91<br>(14) | 93<br>(9) | 92<br>(10) | 92<br>(13) | 89<br>(15) | 90<br>(15) | 88<br>(14) | 84<br>(18) | 85<br>(17) |
| MNN | 3239<br>(1336) | 3335<br>(1163) | 3313<br>(1202) | 5577<br>(1088) | 5792<br>(1676) | 5743<br>(1562) | 7235<br>(1416) | 7783<br>(2150) | 7658<br>(2017) |
| SDN | 1362<br>(709) | 1520<br>(741) | 1484<br>(736) | 1504<br>(582) | 1762<br>(735) | 1703<br>(710) | 1815<br>(586) | 1945<br>(779) | 1916<br>(740) |
| PCN | 93<br>(13) | 94<br>(8) | 94<br>(9) | 92<br>(15) | 93<br>(10) | 93<br>(11) | 88<br>(16) | 86<br>(17) | 86<br>(16) |

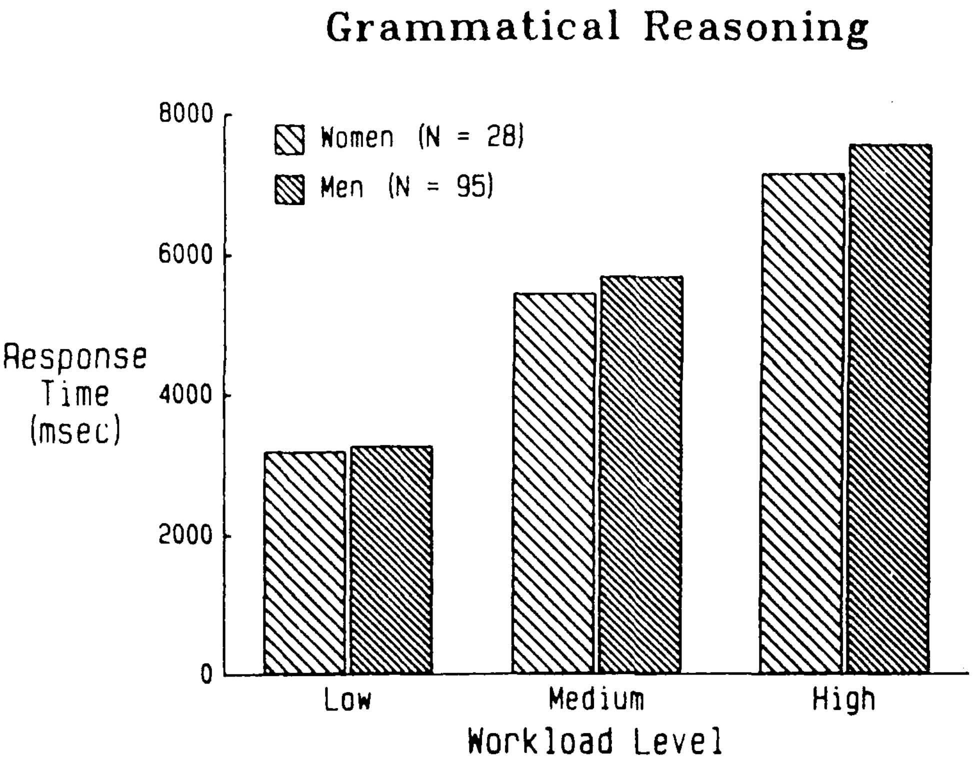


Figure 4-9. Mean Response Time for Grammatical Reasoning - Men vs. Women.

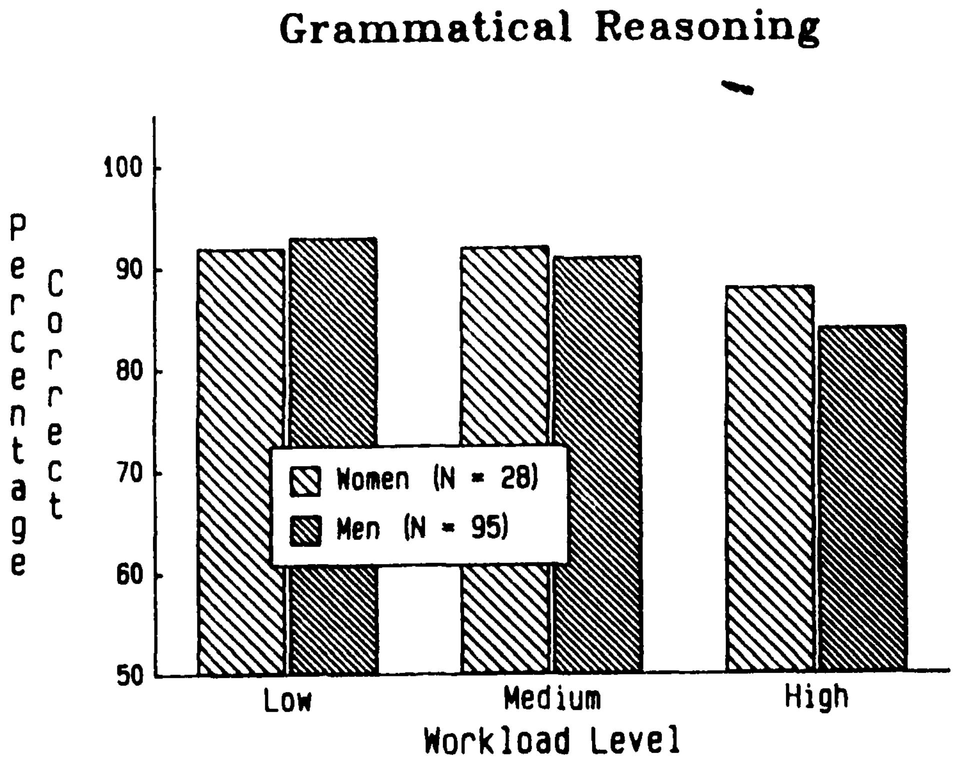


Figure 4-10. Percentage Correct for Grammatical Reasoning - Men vs. Women.

### 4.2.3 Training Data

The means and standard deviations of the major performance measures for Grammatical Reasoning for training Trials 1 through 5 are presented by difficulty level in Table 4-10. Response time and percentage correct are plotted in Figures 4-11 and 4-12. There was significant improvement in both speed and accuracy during training with the largest improvement from Trial 1 to Trial 2. As shown in the figures, asymptotic performance was achieved on different trials depending on the difficulty level and the particular performance measure.

Analysis of variance was used to determine significance between trials for RT and PC using the model presented in Section 4.1.1. A summary of the ANOVA results is presented in Table 4-11. Due to the significant Trial by Level interactions ($p = 0.005$ for RT, $p < 0.0001$ for PC), separate analyses were performed for each level using a reduced model involving only the trial and subject effects. The results of Tukey studentized range tests are summarized in Table 4-12. In all cases, there were no significant differences among Trials 3, 4 and 5 although the improvement trend continued.

Table 4-10. Means (Standard Deviations) of Performance Measures, Grammatical Reasoning - Training Trials.

| Var. | Level | Trial | | | | |
|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 |
| MNO | Low | 3901 (1019) | 3617 (1217) | 3419 (1123) | 3286 (1097) | 3282 (1095) |
| | Med | 6660 (1530) | 6210 (1521) | 5827 (1385) | 5831 (1558) | 5731 (1548) |
| | High | 8706 (1662) | 8295 (1670) | 8112 (1622) | 7857 (2157) | 7737 (1857) |
| SDO | Low | 1742 (656) | 1549 (646) | 1450 (610) | 1442 (637) | 1459 (603) |
| | Med | 1927 (595) | 1655 (571) | 1663 (584) | 1694 (546) | 1720 (629) |
| | High | 2473 (989) | 2406 (941) | 2280 (824) | 2182 (753) | 2164 (825) |
| PCO | Low | 87 (13) | 90 (12) | 91 (10) | 92 ( 9) | 92 ( 9) |
| | Med | 84 (14) | 88 (14) | 89 (12) | 90 (12) | 91 (11) |
| | High | 70 (20) | 74 (18) | 78 (17) | 81 (18) | 83 (16) |
| STIMO | Low | 44 (11) | 50 (13) | 52 (14) | 55 (16) | 55 (15) |
| | Med | 26 ( 6) | 28 ( 7) | 30 ( 7) | 31 (10) | 32 ( 9) |
| | High | 19 ( 4) | 20 ( 5) | 21 ( 5) | 25 (27) | 23 ( 9) |
| MNP | Low | 3850 (1017) | 3493 (1113) | 3347 (1085) | 3257 (1152) | 3211 (1035) |
| | Med | 6612 (1626) | 6140 (1614) | 5733 (1404) | 5786 (1616) | 5647 (1577) |
| | High | 8507 (1587) | 8199 (1838) | 8026 (1904) | 7665 (2243) | 7593 (1972) |
| MNN | Low | 3941 (1121) | 3697 (1308) | 3473 (1233) | 3312 (1119) | 3337 (1167) |
| | Med | 6769 (1661) | 6347 (1697) | 5913 (1487) | 5889 (1593) | 5812 (1579) |
| | High | 8873 (2059) | 8361 (1895) | 8177 (1601) | 8031 (2261) | 7934 (1944) |

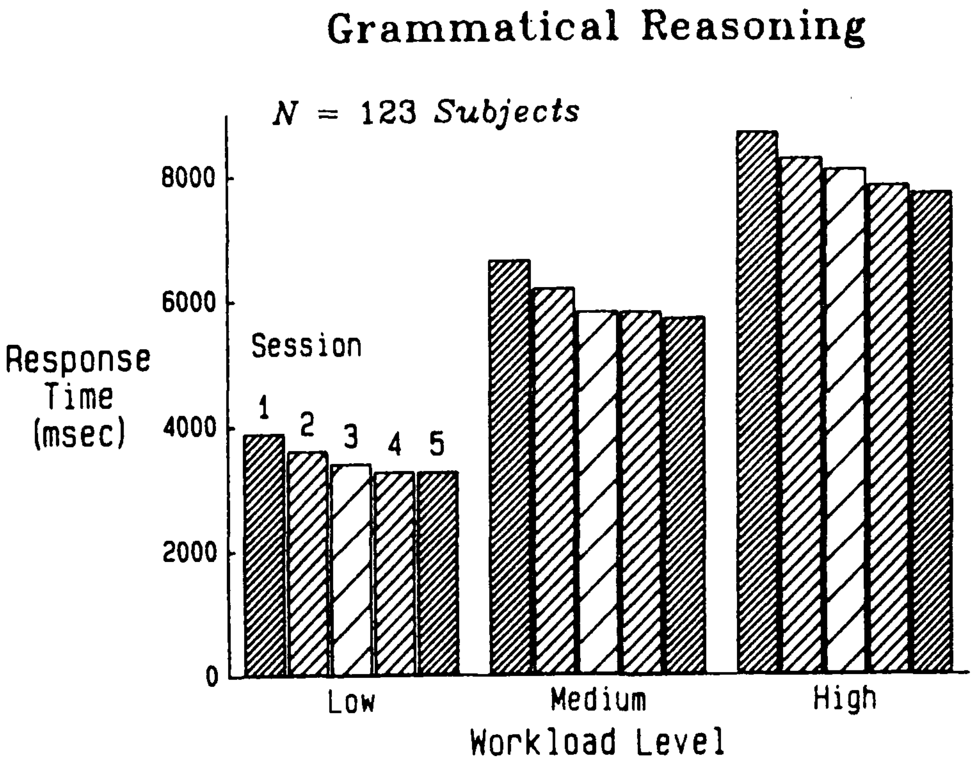


**Figure 4-11. Mean Response Time for Grammatical Reasoning - Trials 1 through 5.**

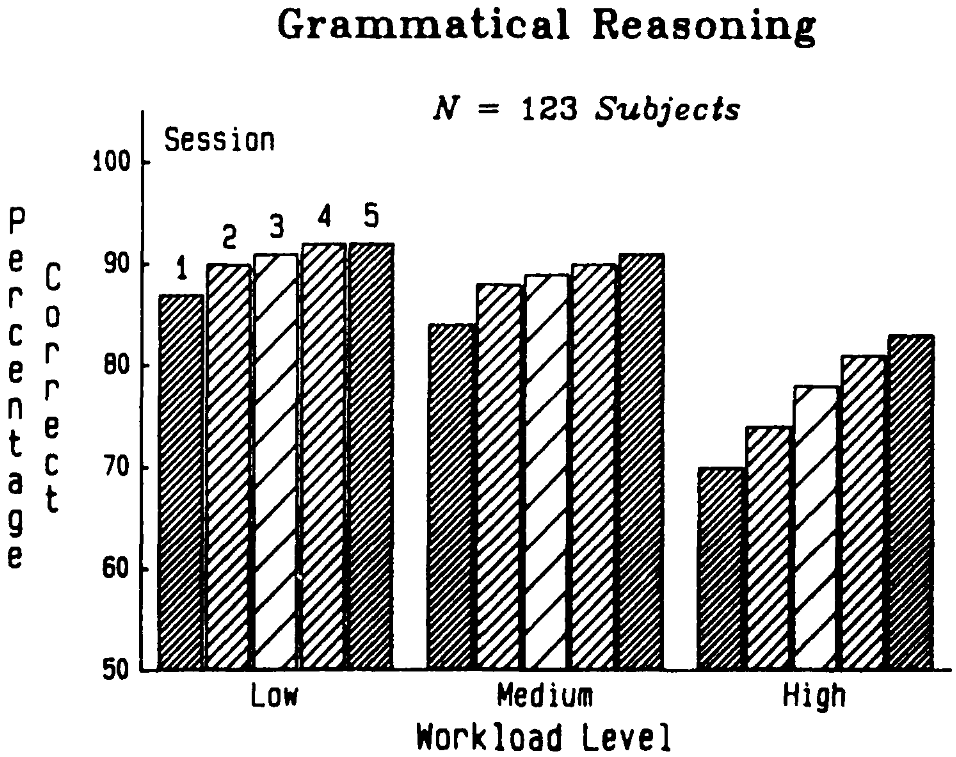


**Figure 4-12. Percentage Correct for Grammatical Reasoning - Trials 1 through 5.**

## Table 4-11. ANOVA Summary for Level and Trial Effects, Grammatical Reasoning - Training Trials.

| Var. | Model | Level | | Trial | | Level * Trial | |
|---|---|---|---|---|---|---|---|
| | $R^2$ | $F_{(2,244)}$ | $p>F$ | $F_{(4,488)}$ | $p>F$ | $F_{(8,976)}$ | $p>F$ |
| GRMNO | 0.96 | 1576.22 | * | 39.90 | * | 2.77 | .0049 |
| GRPCO | 0.85 | 149.26 | * | 50.59 | * | 5.60 | * |

* p < 0.0001

## Table 4-12. Significant (α = .01) Trial Differences by Level.

| Var. | Level | $F_{(4,488)}$ | Trial |
|---|---|---|---|
| MNO | L | 40.74 | $1\ \ 2\ \ \underline{3\ \ 4\ \ 5}$ |
| | M | 30.14 | $1\ \ 2\ \ \underline{4\ \ 3\ \ 5}$ |
| | H | 15.21 | $\underline{1\ \ 2}\ \ \underline{3\ \ 4\ \ 5}$ |
| PCO | L | 21.56 | $1\ \ \underline{2\ \ 3\ \ 5\ \ 4}$ |
| | M | 13.90 | $1\ \ \underline{2\ \ 3\ \ 4\ \ 5}$ |
| | H | 27.15 | $1\ \ \underline{2\ \ 3}\ \ \underline{4\ \ 5}$ |

## 4.3 Interval Production Task

The Interval Production task is the only CTS task that does not provide various difficulty levels. The IPT was originally developed by Michon (1964, 1966) as a secondary task measure of *Perceptual Motor Load.* It was included in the CTS to provide this secondary task capability as part of the test battery. The task measures the subject's ability to produce key taps at a constant interval of approximately 500 msec between taps (two taps per second).

Michon (1966) pointed out that the variance of the tapping sequence was not a suitable measure of tapping irregularity since two vastly different tapping sequences could yield the same interval mean and standard deviation. He proposed a variability score based on the summation of the absolute values of the differences between successive intervals. To adjust for the subject's personal tapping rate, the sum is divided by the mean interval length. This value, adopted as the CTS Variability Score, may be represented as follows:

$$\mathit{CTS\ Variability\ Score} = \frac{\sum |\Delta t|}{\bar{t}} = \frac{\sum |\Delta t|}{T/N}$$

where

$t$ = length of interval between two successive taps,
$T$ = total period of measurement, during which $N$ intervals are produced,
$N$ = number of intervals produced, and
$|\Delta t| = |t_n - t_{n+1}|$.

The assumed logic for the adjustment factor (dividing by $\bar{t}$ ) is that as a person taps faster (shorter $\bar{t}$ ), the $\Delta t$ values are also smaller as is the total summation for a fixed number of taps. Thus, division by $\bar{t}$ standardizes the measure with respect to the tapping rate and also provides a dimensionless quantity. However, a critical oversight is the fact that this measure does not account for differences in the total period of measurement. Michon does not explicitly address this problem. However, in using the task to measure Perceptual Motor Load, Michon computes a PML score reflecting the proportional increase in tapping variability from a baseline condition to a loaded condition using a *fixed number of intervals* under both conditions. A fixed $N$ provides the same number of tapping intervals for comparison and thus division by $\bar{t}$ is permissible.

Unfortunately the CTS IP task is conducted over a fixed time period ($T$) of three minutes during which a varying number of intervals are produced from one trial to the next. Although the same total measurement period exists, differences in tapping rate are not properly accounted for. One approach to this problem is to further adjust the variability score by dividing the original score by the total number of intervals produced. The adjusted variability score becomes $\frac{\sum |\Delta t|}{T}$ which remains a dimensionless quantity. Although it is not obvious from the formula, the adjusted score accounts for differences in tapping rate in two ways. First it accounts for inherently smaller $\Delta t$'s at faster rates and second it accounts for fewer intervals and a smaller sum of $\Delta t$'s at slower rates. Effectively, the tapping rate is removed from the score completely.

### 4.3.1 Interval Production Performance Measures

For the current study, the performance measures selected for presentation and analysis were the mean interval length (IPMN), the standard deviation of interval length (IPSD), the CTS variability score provided by the CTS software (IPVS1), and the adjusted variability score defined above (IPVS2). These measures are summarized by gender and trial in Table 4-13 and Figures 4-13 through 4-16. As mentioned in Section 3.4, all summaries and analyses were based on 120 subjects since the variability scores for one female subject (#6) and two male subjects (#73, #76) were 5 to 9 standard deviations above the mean. Univariate summaries of the performance measures based on the Trial 6 and Trial 8 data for the 120 subjects are provided in Appendix A-3.

The mean interval duration for Trials 6 and 8 combined was 506 msec ($s$ = 129 msec) indicating that on the average subjects were extremely close to the desired tapping rate of two taps per second. The standard deviations of the tapping intervals produced during the three-minute trials averaged 51 msec and were quite constant across trials and genders. The CTS Variability Score averaged 28.6 ($s$ = 13.5) while the Adjusted Variability Score averaged 0.0767 ($s$ = 0.0314).

The Trial 6 - Trial 8 correlation for the mean interval length was 0.82 verifying the concept that each subject develops a fairly constant yet personal tapping rate. However, the correlations for the other performance measures were quite low (0.56 for IPSD, 0.57 for IPVS1, and 0.49 for IPVS2) indicating overall poor stability for this task.

Table 4-13. Means (Standard Deviations) of Performance Measures, Interval Production - Baseline Trials.

| Gender | Females | | | Males | | |
|---|---|---|---|---|---|---|
| Trial | 06 | 08 | Both | 06 | 08 | Both |
| IPMN | 515<br>(104) | 520<br>(114) | 517<br>(108) | 512<br>(149) | 494<br>(119) | 503<br>(135) |
| IPSD | 52<br>(25) | 49<br>(41) | 50<br>(33) | 52<br>(43) | 52<br>(41) | 52<br>(42) |
| IPVS1 | 27.6<br>(13.4) | 26.6<br>(16.8) | 27.1<br>(15.1) | 28.1<br>(10.5) | 30.0<br>(15.1) | 29.1<br>(13.0) |
| IPVS2 | .0755<br>(.0318) | .0748<br>(.0418) | .0751<br>(.0368) | .0760<br>(.0255) | .0783<br>(.0336) | .0771<br>(.0298) |

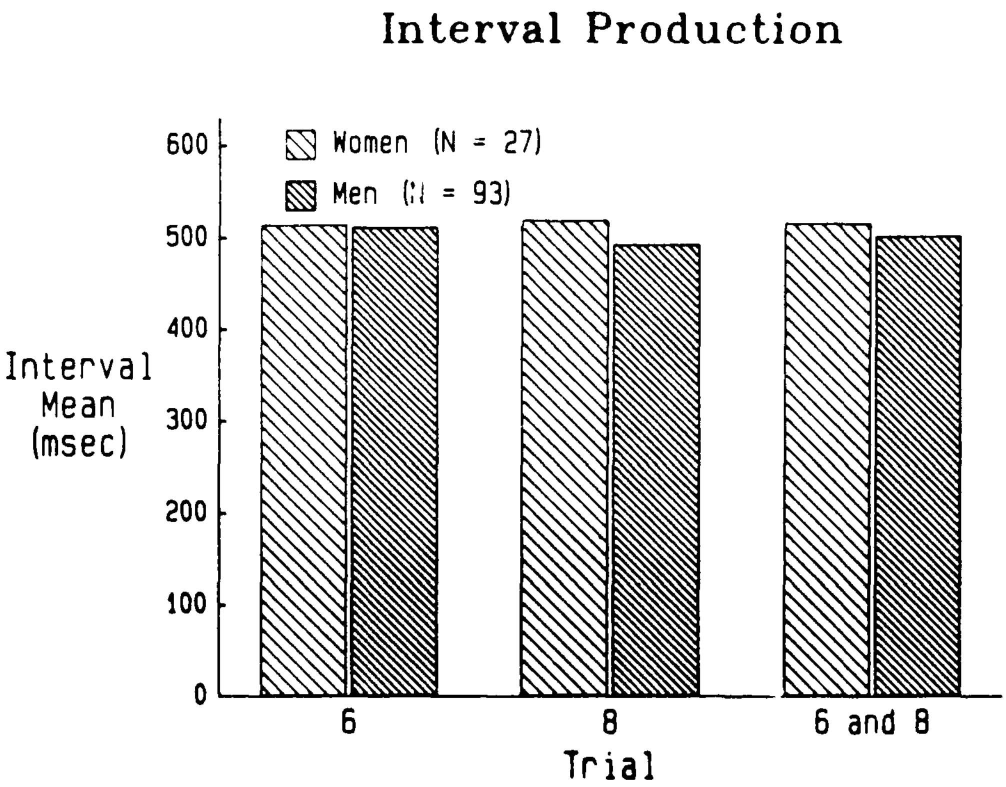


Figure 4-13. Interval Mean for Interval Production - Baseline Trials.

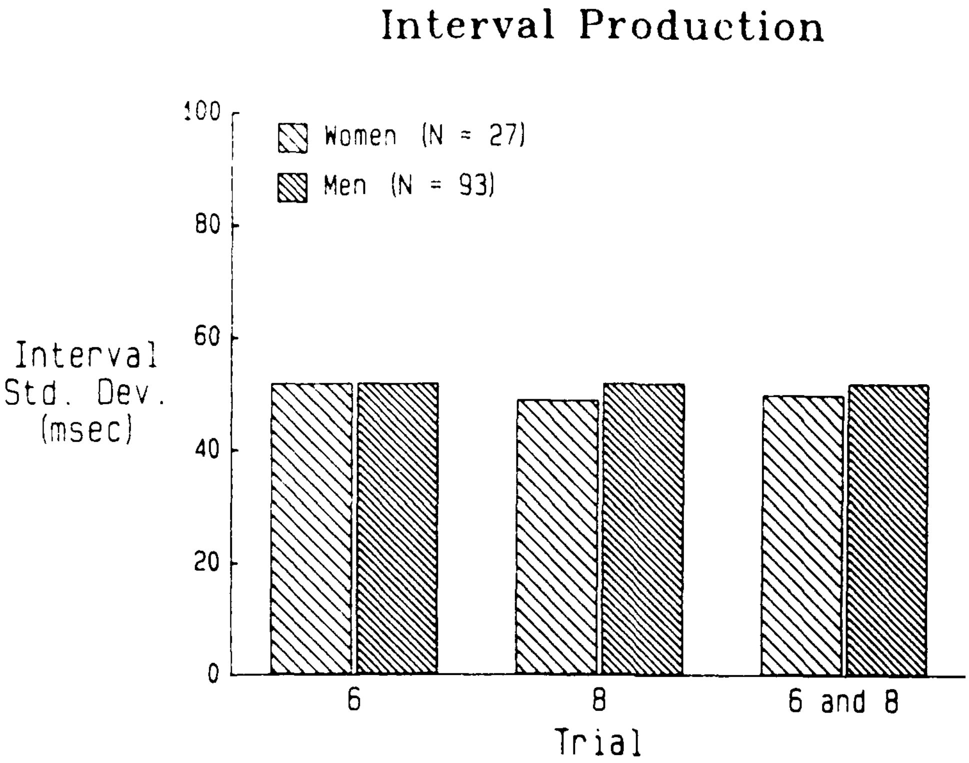


Figure 4-14. Interval Standard Deviation for Interval Production - Baseline Trials.

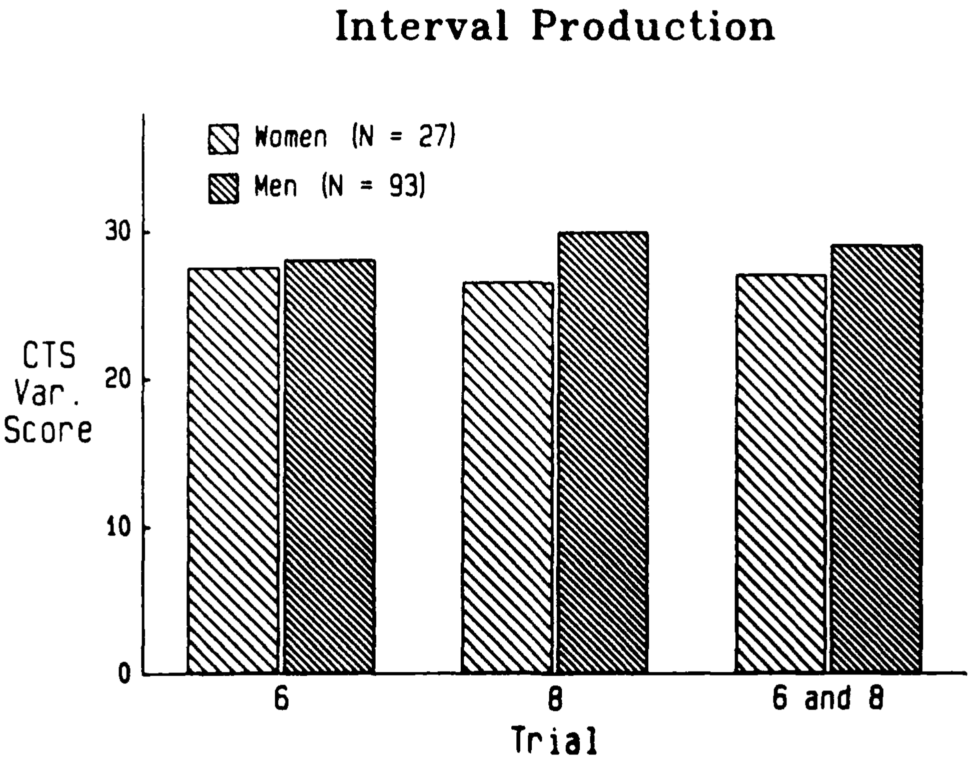


Figure 4-15. CTS Variability Score for Interval Production - Baseline Trials.

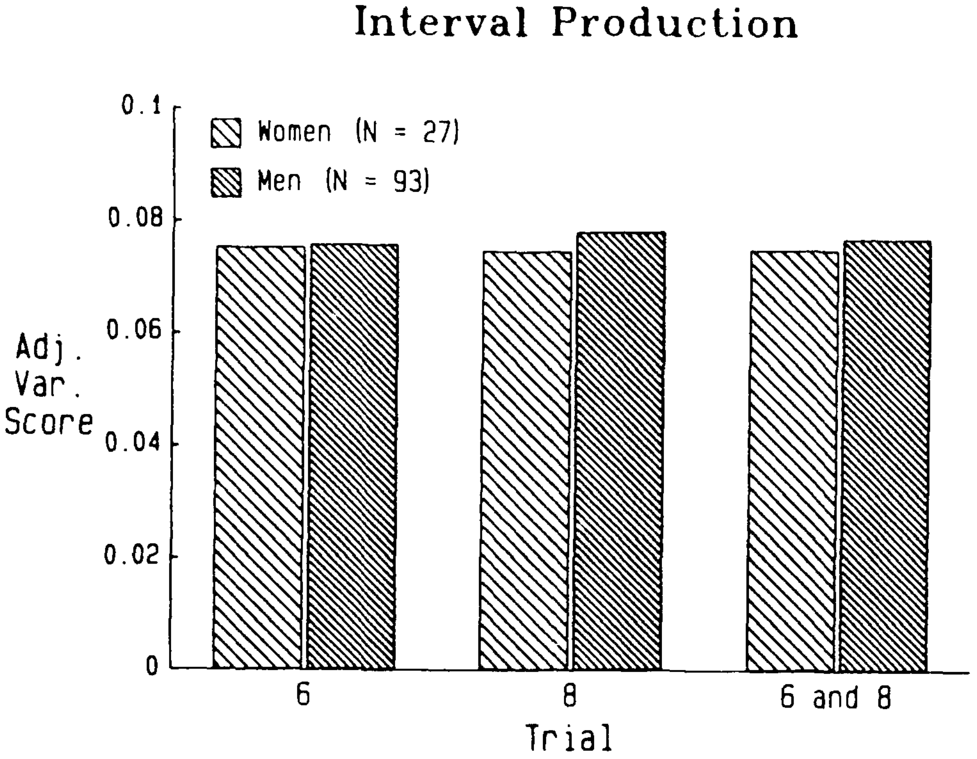


**Figure 4-16. Adjusted Variability Score for Interval Production - Baseline Trials.**

### 4.3.2 Trial, Gender and Prototype Analyses

Differences between Trials 6 and 8 were examined using an additive two-way ANOVA model involving Trial and Subject. The results are summarized in Table 4-14. A marginally significant difference (p = 0.076) existed only for the mean interval duration with Trial 8 (500 msec) having a slightly faster mean tapping rate than Trial 6 (513 msec). There were no trial differences for any of the variables which measured tapping variability.

There were no significant differences between men and women for any of the performance measures. Likewise, there were no significant differences among the various prototypes although the Effort prototype group tended to produce shorter intervals (faster tapping) with greater variability than the other groups.

**Table 4-14. ANOVA Summary for Trial Effect,
Interval Production - Baseline Trials.**

**Interval Mean** ($R^2$ = 0.90)

| Source | DF | Sum of Squares (x1000) | Mean Square (x1000) | F | p > F |
|---|---|---|---|---|---|
| Trial | 1 | 11 | 11 | 3.20 | 0.0763 |
| Subject | 119 | 3591 | 30 | 9.18 | 0.0001 |
| Error | 119 | 391 | 3 | | |

**Interval Standard Deviation** ($R^2$ = 0.78)

| Source | DF | Sum of Squares (x100) | Mean Square (x100) | F | p > F |
|---|---|---|---|---|---|
| Trial | 1 | 1 | 1 | 0.13 | 0.7173 |
| Subject | 119 | 3029 | 25 | 3.59 | 0.0001 |
| Error | 119 | 844 | 7 | | |

**CTS Variability Score** ($R^2$ = 0.77)

| Source | DF | Sum of Squares | Mean Square | F | p > F |
|---|---|---|---|---|---|
| Trial | 1 | 95 | 95 | 1.14 | 0.2884 |
| Subject | 119 | 33501 | 282 | 3.36 | 0.0001 |
| Error | 119 | 9978 | 84 | | |

**Adjusted Variability Score** ($R^2$ = 0.74)

| Source | DF | Sum of Squares (x0.0001) | Mean Square (x0.0001) | F | p > F |
|---|---|---|---|---|---|
| Trial | 1 | 1 | 1 | 0.28 | 0.5956 |
| Subject | 119 | 1737 | 15 | 2.81 | 0.0001 |
| Error | 119 | 2357 | 20 | | |

### 4.3.3 Training Data

The means and standard deviations of the performance measures for Interval Production for training Trials 1 through 5 and baseline Trials 6 and 8 are presented in Table 4-15. The performance measures are plotted in Figures 4-17 and 4-18. Note that *Standard Deviation* in Figure 4-17 represents the average of the standard deviations of the intervals produced during the three-minute trials rather than the standard deviation across subjects associated with the mean interval length. For all variables, there were negligible changes in performance during training. In fact, the lowest variability scores occurred on Trial 1. This is possibly due to the perceived simplicity of the task and the accompanying lack of concentration by the subjects beyond Trial 1. A summary of analysis of variance and Tukey studentized range tests for the five training trials is presented in Table 4-16.

**Table 4-15. Means (Standard Deviations) of Performance Measures, Interval Production - Training and Baseline Trials.**

| Var. | Trial | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 8 |
| IPMN | 529 (149) | 494 (120) | 499 (135) | 496 (125) | 517 (152) | 513 (140) | 500 (118) |
| IPSD | 59 (52) | 52 (47) | 54 (42) | 56 (42) | 55 (70) | 52 (40) | 51 (41) |
| IPVS1 | 28 (16) | 37 (42) | 35 (31) | 36 (33) | 30 (24) | 28 (11) | 29 (16) |
| IPVS2* | 777 (318) | 897 (806) | 871 (573) | 920 (651) | 800 (553) | 759 (269) | 775 (354) |

* times a scale factor of 0.0001

**Table 4-16. ANOVA and Tukey (α = .01) Summary for Trial Effects, Interval Production - Training Trials.**

| Var. | Model $R^2$ | $F_{(4,476)}$ | $p > F$ | Trial | | | | |
|---|---|---|---|---|---|---|---|---|
| IPMN | 0.75 | 4.61 | .0012 | 1 | 5 | 3 | 4 | 2 |
| IPSD | 0.48 | 0.50 | .7332 | 1 | 4 | 5 | 3 | 2 |
| IPVS1 | 0.41 | 2.58 | .0369 | 2 | 4 | 3 | 5 | 1 |
| IPVS2 | 0.42 | 1.75 | .1387 | 4 | 2 | 3 | 5 | 1 |

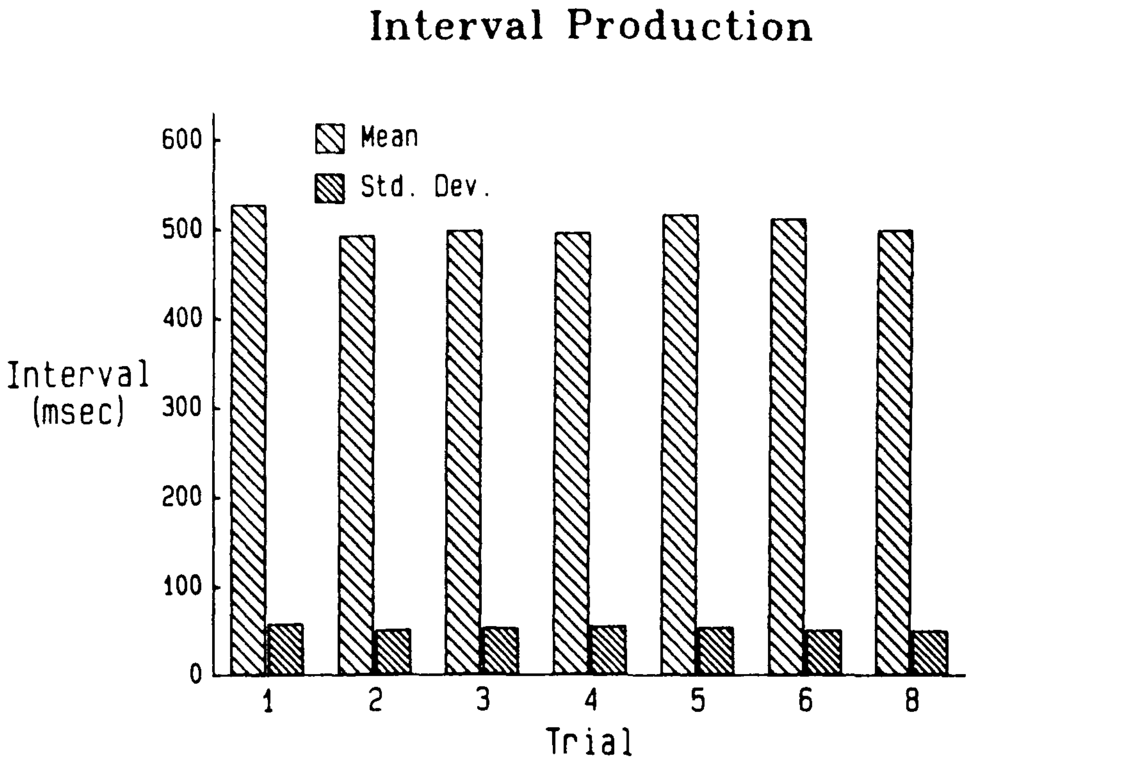


Figure 4-17. Interval Statistics for Interval Production - Trials 1 through 8.

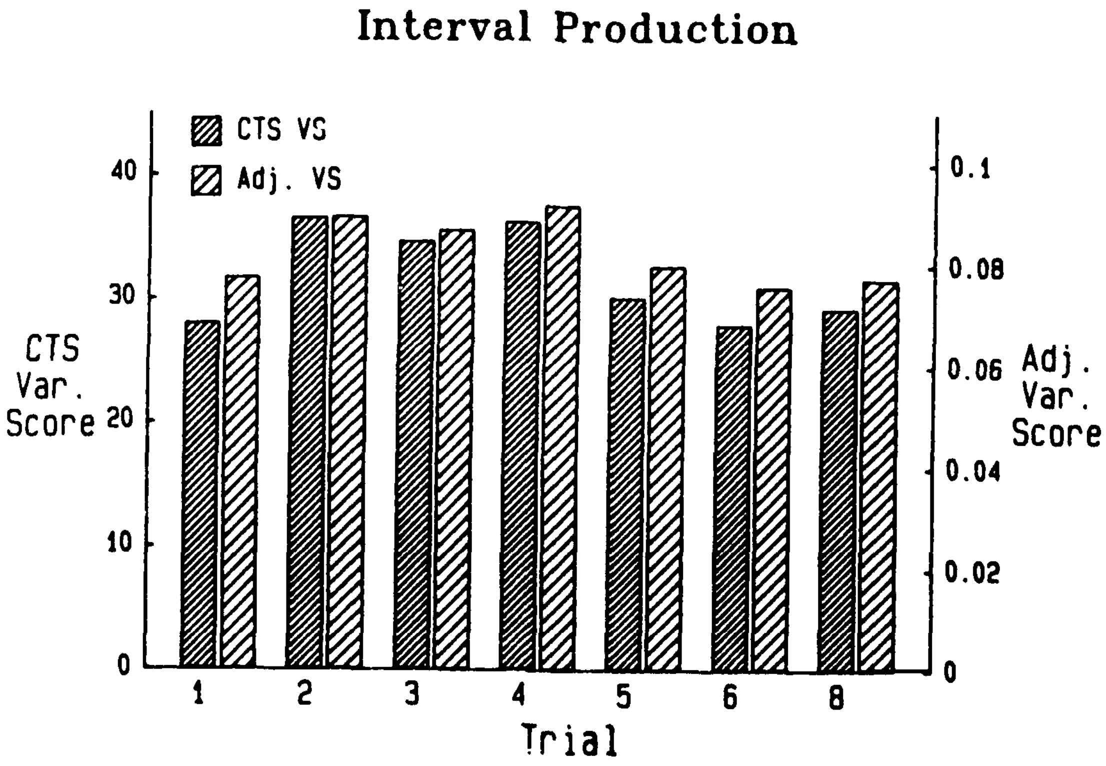


Figure 4-18. Variability Scores for Interval Production - Trials 1 through 8.

## 4.4 Linguistic Processing Task

The means and standard deviations for the Linguistic Processing performance measures are presented in Table 4-17 for Trials 6 and 8. Overall mean response time (RT) and percentage correct (PC) are presented in Figures 4-19 and 4-20 with the intertrial correlations represented by the *r* values in the figures. Univariate summaries of the overall response time and proportion correct measures for the three difficulty levels are provided in Appendix A-4. Data for these summaries included the Trial 6 and Trial 8 data for all 123 subjects.

Response times were reasonably fast for this task and were approximately 0.5, 0.8 and 1.6 seconds for the low, medium and high difficulty levels respectively. The larger performance difference between the medium and high levels of LP indicated a greater change in task difficulty between these levels. The standard deviation of response time increased substantially with increasing difficulty from approximately 0.2 to 0.7 seconds. Proportion correct was very high for the low (0.97) and medium (0.96) levels but dropped off slightly for the high level (0.90) with an accompanying increase in standard deviation from 0.04 to 0.07.

As with CR and GR, response times were faster (by 50 to 100 msec) for the *MATCH* (right button) responses compared with the *NON-MATCH* (left button) responses. However, proportion correct was approximately the same for both types of responses at the low (Physical ID) and medium (Category Match) levels. It was slightly higher for *MATCH* responses at the low level and slightly lower for *NON-MATCH* responses at the medium level. In contrast, the proportion correct was substantially lower (0.84 vs. 0.96) for *MATCH* responses at the high (Antonym) level. This indicates that subjects made fewer errors classifying words that were unrelated than they did identifying words that were in fact antonyms. In other words, a higher proportion of words that *were* antonyms were considered unrelated.

The *r* values for response time were reasonably large for the medium (0.80) and high (0.87) difficulty levels but somewhat smaller (0.69) for the low levels. This was probably due to the fast response times achieved by virtually all subjects at the low level of the Linguistic Processing task. This same phenomenon (easy task with high accuracy) was probably responsible for the very low *r* values (0.48, 0.41) for proportion correct at the low and medium levels. The harder Antonym task provided an intertrial correlation of 0.79.

## Table 4-17. Means (Standard Deviations) of Performance Measures, Linguistic Processing - Baseline Trials.

| Level | Low | | | Medium | | | High | | |
|---|---|---|---|---|---|---|---|---|---|
| Trial | 06 | 08 | Both | 06 | 08 | Both | 06 | 08 | Both |
| MNO | 530<br>(117) | 517<br>(101) | 523<br>(109) | 812<br>(240) | 772<br>(257) | 792<br>(249) | 1600<br>(441) | 1557<br>(459) | 1578<br>(450) |
| SDO | 178<br>(165) | 188<br>(211) | 183<br>(189) | 365<br>(286) | 330<br>(296) | 348<br>(291) | 748<br>(384) | 739<br>(418) | 743<br>(400) |
| PCO | 97<br>(4) | 97<br>(4) | 97<br>(4) | 96<br>(4) | 96<br>(3) | 96<br>(3) | 90<br>(7) | 90<br>(8) | 90<br>(7) |
| STIMO | 249<br>(35) | 252<br>(35) | 250<br>(35) | 183<br>(37) | 191<br>(41) | 187<br>(39) | 101<br>(21) | 104<br>(24) | 103<br>(23) |
| MNP | 510<br>(106) | 496<br>(87) | 503<br>(97) | 754<br>(209) | 721<br>(239) | 737<br>(225) | 1569<br>(414) | 1556<br>(550) | 1563<br>(486) |
| SDP | 156<br>(180) | 151<br>(156) | 154<br>(168) | 318<br>(250) | 272<br>(259) | 295<br>(255) | 726<br>(359) | 754<br>(519) | 740<br>(446) |
| PCP | 97<br>(3) | 97<br>(3) | 97<br>(3) | 95<br>(6) | 95<br>(4) | 95<br>(5) | 85<br>(13) | 84<br>(14) | 84<br>(14) |
| MNN | 555<br>(146) | 541<br>(128) | 548<br>(137) | 875<br>(281) | 825<br>(293) | 850<br>(287) | 1645<br>(554) | 1579<br>(479) | 1612<br>(518) |
| SDN | 185<br>(156) | 207<br>(270) | 196<br>(221) | 387<br>(332) | 354<br>(335) | 371<br>(333) | 729<br>(444) | 694<br>(408) | 711<br>(426) |
| PCN | 96<br>(8) | 96<br>(6) | 96<br>(7) | 97<br>(3) | 97<br>(2) | 97<br>(3) | 96<br>(4) | 96<br>(4) | 96<br>(4) |

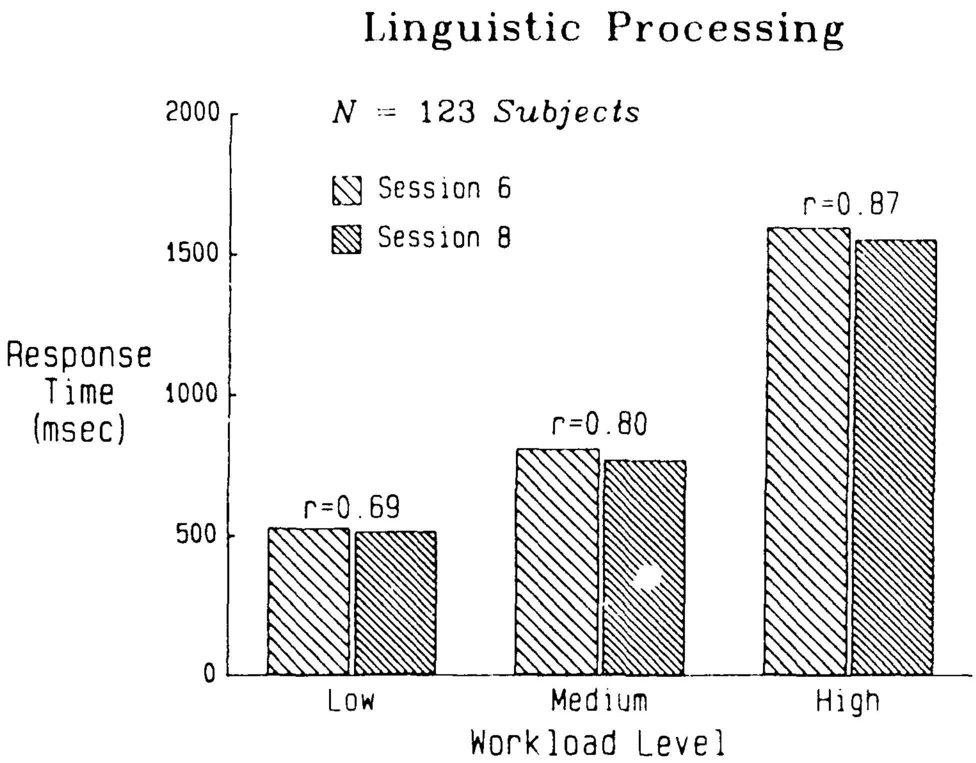


Figure 4-19. Mean Response Time for Linguistic Processing - Trials 6 and 8.

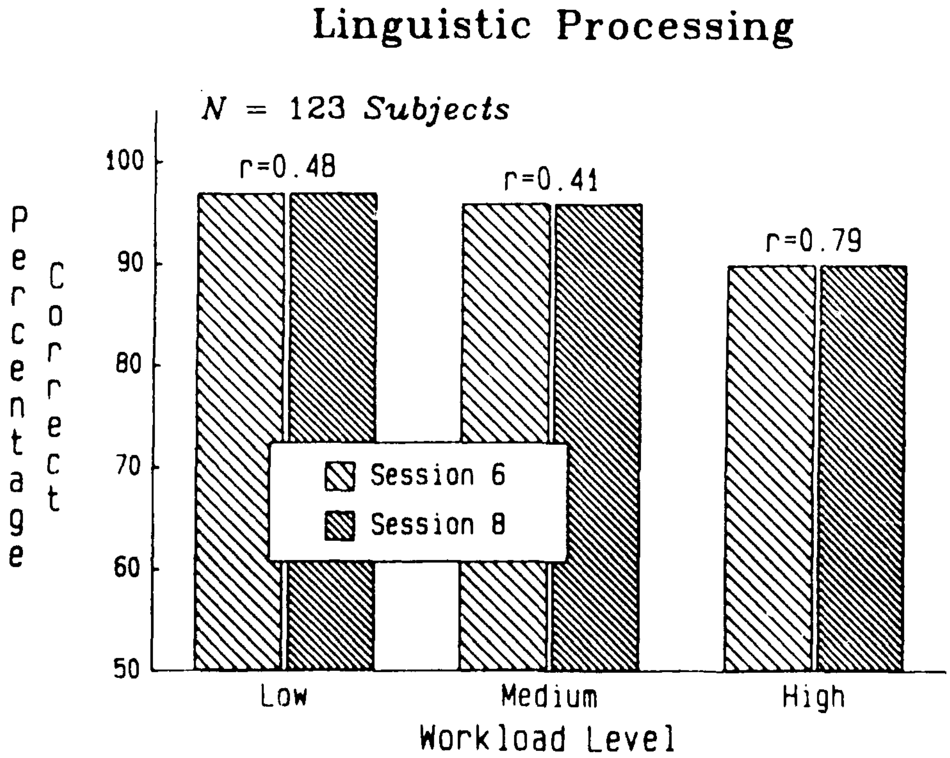


Figure 4-20. Percentage Correct for Linguistic Processing - Trials 6 and 8.

### 4.4.1 Level and Trial Analyses

Analysis of variance was performed to verify the difficulty level manipulation and examine any trial differences using the model presented in Section 4.1.1. The results of the analyses for response time and proportion correct are summarized in Table 4-18. The model $R^2$ values were 0.99 for RT and 0.92 for PC. The Tukey studentized range test at $\alpha = 0.01$ demonstrated that all three difficulty levels differed significantly for RT. However, as with GR, the low and medium levels did not differ with respect to PC as both of these levels represent fairly easy tasks. The mean response time for Trial 8 was significantly lower (p = 0.002) than the RT for Trial 6 with no Level by Trial interaction. However, the difference averaged less than 32 msec. The proportion correct values were identical for Trials 6 and 8 at all levels.

**Table 4-18. ANOVA Summary for Level and Trial Effects, Linguistic Processing - Baseline Trials.**

**Response Time**

| Source | DF | Sum of Squares (x1000) | Mean Square (x1000) | F | p > F |
|---|---|---|---|---|---|
| Level (L) | 2 | 147767 | 73883 | 721.66 | 0.0001 |
| Trial (T) | 1 | 188 | 188 | 9.65 | 0.0024 |
| L by T | 2 | 36 | 18 | 1.48 | 0.2300 |
| Subject (S) | 122 | 37125 | 304 | 25.07 | 0.0001 |
| L by S | 244 | 24981 | 102 | 8.44 | 0.0001 |
| T by S | 122 | 2380 | 20 | 1.61 | 0.0009 |
| Error | 244 | 2961 | 12 | | |

**Proportion Correct**

| Source | DF | Sum of Squares (x0.001) | Mean Square (x0.001) | F | p > F |
|---|---|---|---|---|---|
| Level (L) | 2 | 699 | 350 | 99.80 | 0.0001 |
| Trial (T) | 1 | 0 | 0 | 0.00 | 0.9715 |
| L by T | 2 | 1 | 1 | 0.71 | 0.4950 |
| Subject (S) | 122 | 724 | 6 | 7.12 | 0.0001 |
| L by S | 244 | 855 | 4 | 4.20 | 0.0001 |
| T by S | 122 | 104 | 1 | 1.02 | 0.4375 |
| Error | 244 | 203 | 1 | | |

### 4.4.2 Gender and Prototype Analyses

As described in Section 4.1.2 for CR, further analyses of gender and prototype differences were performed. The performance measures are presented separately for men and women in Table 4-19 and Figures 4-21 and 4-22. There were no differences in response time or proportion correct between men and women and no significant interactions involving gender.

As with CR and GR, the Effort prototype group tended to have faster response times than the other groups but the difference was not significant. There was a marginally significant ($p = 0.06$) Trial by Prototype interaction for RT. There were no differences for the proportion correct measure. Refer to Section 5.2 for a discussion of the prototype grouping.

### 4.4.3 Training Data

The means and standard deviations of the major performance measures for Linguistic Processing for training Trials 1 through 5 are presented by difficulty level in Table 4-20. Response time and percentage correct are plotted in Figures 4-23 and 4-24. Analysis of variance was used to determine significance between trials for RT and PC using the model presented in Section 4.1.1. A summary of the ANOVA results is presented in Table 4-21. There was significant improvement in speed but only marginally significant ($p = 0.05$) improvement in accuracy during training. For RT, there was a highly significant Trial by Level interaction with greater improvement at the higher difficulty levels. From a practical standpoint, there was no improvement in accuracy (less than 1%) for any difficulty level from the first to the fifth trial.

Due to the significant Trial by Level interactions ($p < 0.0001$ for RT, $p = 0.0003$ for PC), separate analyses were performed for each level using a reduced model involving only the trial and subject effects. The results of Tukey studentized range tests for the individual analyses are summarized in Table 4-22. With one exception, there were no significant differences among Trials 3, 4 and 5 although the improvement trend continued for RT. The exception involved RT at the medium level where Trial 3 differed from Trials 4 and 5.

**Table 4-19. Means (Standard Deviations) of Performance Measures by Gender, Linguistic Processing - Baseline Trials.**

| Level | Low | | | Medium | | | High | | |
|---|---|---|---|---|---|---|---|---|---|
| Gender | Fem | Male | Both | Fem | Male | Both | Fem | Male | Both |
| MNO | 524<br>(127) | 523<br>(103) | 523<br>(109) | 764<br>(246) | 800<br>(250) | 792<br>(249) | 1507<br>(396) | 1599<br>(463) | 1578<br>(450) |
| SDO | 179<br>(189) | 184<br>(189) | 183<br>(189) | 318<br>(271) | 356<br>(297) | 348<br>(291) | 714<br>(357) | 752<br>(413) | 743<br>(400) |
| PCO | 98<br>(3) | 96<br>(4) | 97<br>(4) | 97<br>(5) | 96<br>(3) | 96<br>(3) | 90<br>(9) | 90<br>(7) | 90<br>(7) |
| STIMO | 247<br>(39) | 251<br>(34) | 250<br>(35) | 190<br>(36) | 187<br>(40) | 187<br>(39) | 105<br>(21) | 102<br>(23) | 103<br>(23) |
| MNP | 507<br>(121) | 502<br>(89) | 503<br>(97) | 706<br>(229) | 747<br>(223) | 737<br>(225) | 1531<br>(572) | 1572<br>(458) | 1563<br>(486) |
| SDP | 160<br>(211) | 152<br>(154) | 154<br>(168) | 235<br>(167) | 313<br>(273) | 295<br>(255) | 752<br>(513) | 736<br>(425) | 740<br>(446) |
| PCP | 98<br>(2) | 97<br>(3) | 97<br>(3) | 96<br>(8) | 95<br>(4) | 95<br>(5) | 84<br>(17) | 84<br>(12) | 84<br>(14) |
| MNN | 544<br>(147) | 549<br>(135) | 548<br>(137) | 826<br>(286) | 856<br>(288) | 850<br>(287) | 1530<br>(440) | 1636<br>(537) | 1612<br>(518) |
| SDN | 177<br>(175) | 202<br>(233) | 196<br>(221) | 355<br>(344) | 376<br>(331) | 371<br>(333) | 653<br>(377) | 729<br>(438) | 711<br>(426) |
| PCN | 97<br>(6) | 96<br>(7) | 96<br>(7) | 98<br>(2) | 97<br>(3) | 97<br>(3) | 96<br>(4) | 96<br>(4) | 96<br>(4) |

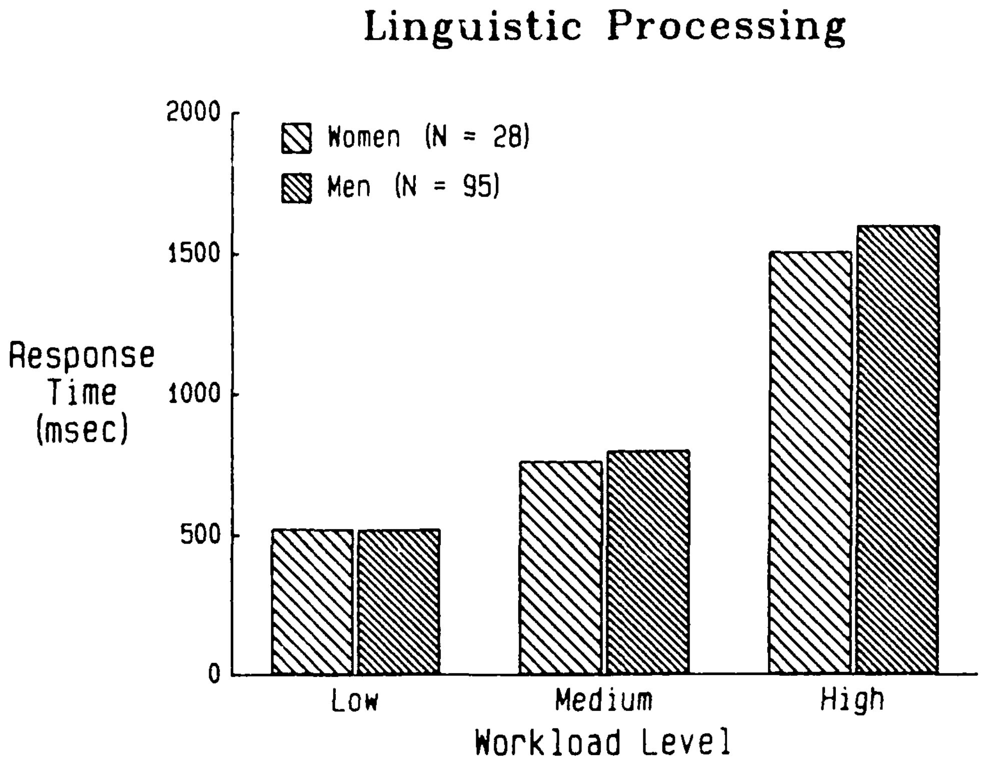


Figure 4-21. Mean Response Time for Linguistic Processing - Men vs. Women.

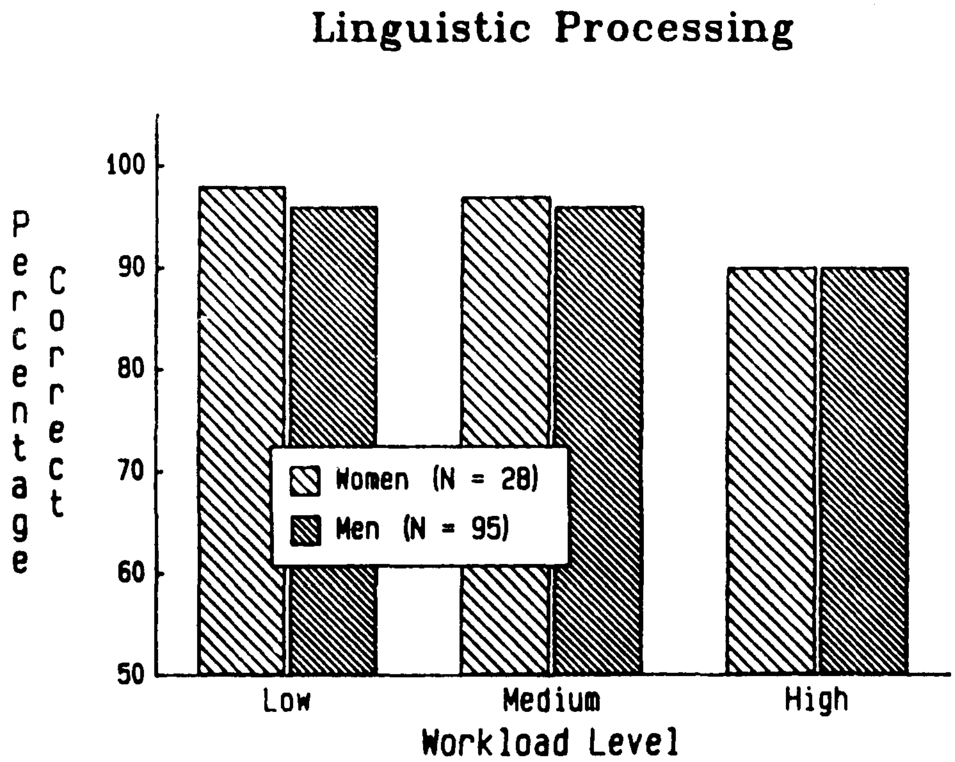


Figure 4-22. Percentage Correct for Linguistic Processing - Men vs. Women.

Table 4-20. Means (Standard Deviations) of Performance Measures, Linguistic Processing - Training Trials.

| Var. | Level | Trial | | | | |
|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 |
| MNO | Low | 650 (127) | 591 (108) | 559 (107) | 538 (103) | 542 (142) |
| | Med | 1195 (350) | 997 (272) | 926 (286) | 854 (231) | 848 (251) |
| | High | 2129 (584) | 1825 (511) | 1705 (469) | 1625 (441) | 1643 (521) |
| SDO | Low | 291 (193) | 239 (180) | 202 (141) | 172 (115) | 186 (212) |
| | Med | 622 (337) | 475 (278) | 415 (301) | 367 (225) | 378 (330) |
| | High | 1031 (438) | 838 (429) | 771 (393) | 736 (375) | 753 (436) |
| PCO | Low | 96 (8) | 97 (7) | 96 (7) | 97 (5) | 97 (5) |
| | Med | 94 (7) | 95 (6) | 96 (4) | 96 (4) | 96 (4) |
| | High | 91 (6) | 91 (6) | 90 (7) | 90 (7) | 91 (7) |
| STIMO | Low | 209 (32) | 225 (30) | 236 (30) | 244 (31) | 243 (36) |
| | Med | 132 (28) | 153 (31) | 164 (34) | 174 (33) | 176 (36) |
| | High | 77 (16) | 89 (19) | 95 (20) | 100 (22) | 99 (22) |
| MNP | Low | 599 (102) | 549 ( 86) | 529 ( 88) | 507 ( 75) | 519 (133) |
| | Med | 1083 (316) | 919 (253) | 846 (234) | 788 (208) | 780 (207) |
| | High | 2074 (590) | 1814 (573) | 1719 (483) | 1624 (475) | 1665 (551) |
| MNN | Low | 710 (176) | 635 (141) | 595 (150) | 572 (145) | 568 (166) |
| | Med | 1306 (399) | 1079 (312) | 1006 (342) | 920 (262) | 918 (309) |
| | High | 2191 (676) | 1864 (594) | 1705 (502) | 1633 (483) | 1636 (545) |

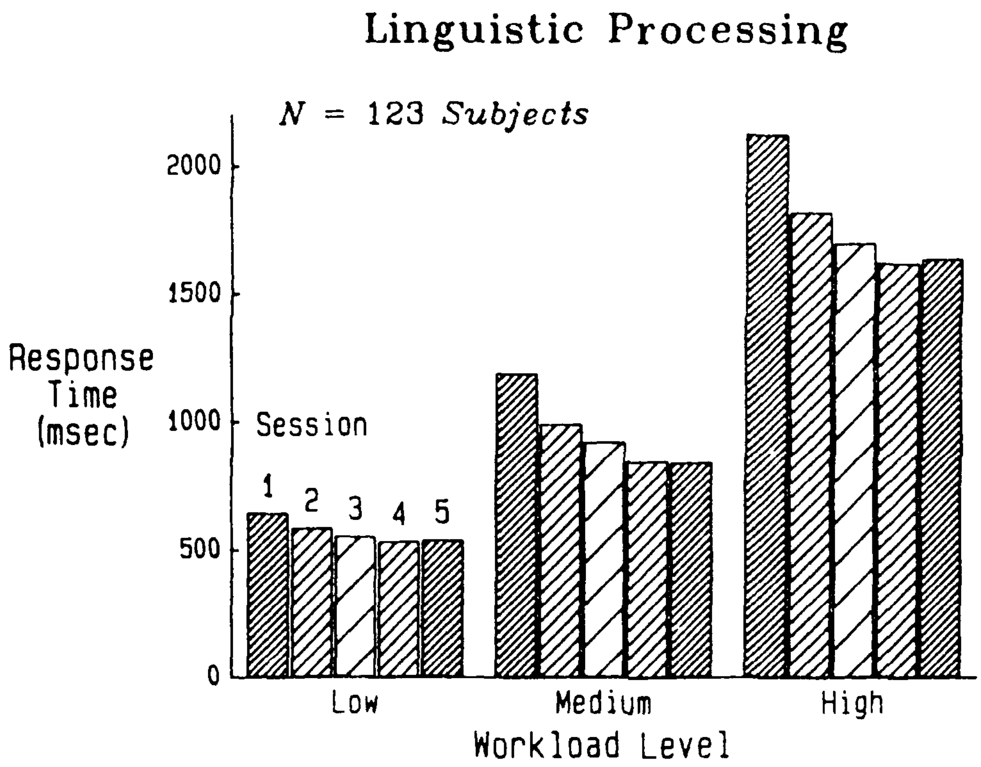


**Figure 4-23. Mean Response Time for Linguistic Processing - Trials 1 through 5.**

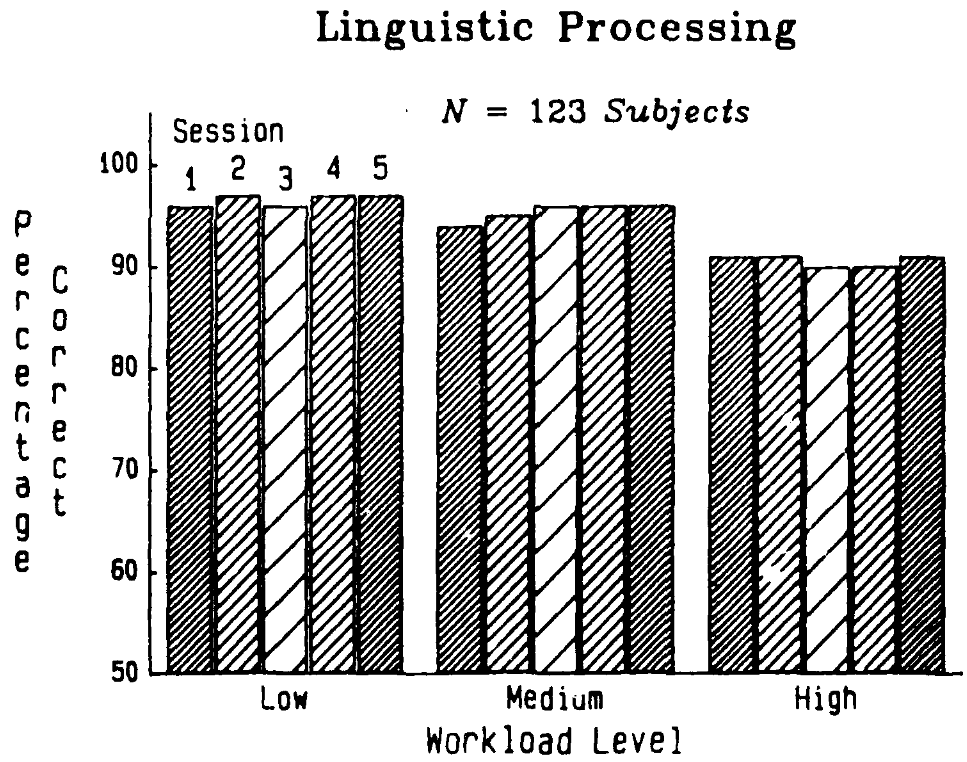


**Figure 4-24. Percentage Correct for Linguistic Processing - Trials 1 through 5.**

## Table 4-21. ANOVA Summary for Level and Trial Effects, Linguistic Processing - Training Trials.

| Var. | Model | Level | | Trial | | Level * Trial | |
|---|---|---|---|---|---|---|---|
| | $R^2$ | $F_{(2,244)}$ | $p>F$ | $F_{(4,488)}$ | $p>F$ | $F_{(8,976)}$ | $p>F$ |
| LPMNO | 0.97 | 768.16 | * | 139.52 | * | 41.09 | * |
| LPPCO | 0.86 | 48.37 | * | 2.38 | .0510 | 3.64 | .0003 |

* p < 0.0001

## Table 4-22. Significant (α = .01) Trial Differences by Level.

| Var. | Level | $F_{(4,488)}$ | Trial | | | | |
|---|---|---|---|---|---|---|---|
| | L | 46.82 | 1 | 2 | 3 | 5 | 4 |
| MNO | M | 146.73 | 1 | 2 | 3 | 4 | 5 |
| | H | 83.83 | 1 | 2 | 3 | 5 | 4 |
| | L | 1.85 | 1 | 3 | 4 | 2 | 5 |
| PCO | M | 7.01 | 1 | 2 | 3 | 4 | 5 |
| | H | 0.82 | 1 | 2 | 5 | 3 | 4 |

## 4.5 Mathematical Processing Task

The means and standard deviations for the Mathematical Processing performance measures are presented in Table 4-23 for Trials 6 and 8. Overall mean response time (RT) and percentage correct (PC) are presented in Figures 4-25 and 4-26 with the intertrial correlations represented by the *r* values in the figures. Univariate summaries of the overall response time and proportion correct measures for the three difficulty levels are provided in Appendix A-5. Data for these summaries included the Trial 6 and Trial 8 data for all 123 subjects.

Response times for MP were approximately 0.5, 1.5 and 2.5 seconds for the low, medium and high difficulty levels with standard deviations that increased substantially with increasing difficulty from approximately 0.2 to 1.0 seconds. Proportion correct was an extremely stable 0.97 for all levels with a slight increase in standard deviation (from 0.03 to 0.05) at the high level only.

In contrast with the other central processing tasks, both response time and proportion correct were essentially equal for the right-button ( > 5) and left-button ( < 5) responses at the low and medium levels. At the high level, however, right-button responses were slightly faster. This is possibly due to the fact that stimuli involving only addition operations would typically lead to larger sums and allow a *greater than 5* decision to be made sooner.

The *r* values for response time were large (0.74 to 0.91), indicating high stability in performance for this task. The values for percentage correct were extremely low (0.17 to 0.60) probably due to the consistently high accuracy achieved by most subjects at all levels.

### 4.5.1 Level and Trial Analyses

Analysis of variance was performed to verify the difficulty level manipulation and examine any trial differences using the model presented in Section 4.1.1. The results of the analyses for response time and proportion correct are summarized in Table 4-24. The model $R^2$ values were 0.99 for RT and 0.85 for PC. The Tukey studentized range test at $\alpha = 0.01$ demonstrated that all three difficulty levels differed significantly for RT. There were no differences among levels for PC. The mean response time for Trial 8 was significantly lower (p = 0.008) than the RT for Trial 6 with a slightly significant Level by Trial interaction due to larger trial differences with increasing difficulty level. The proportion correct values were identical for Trials 6 and 8.

Table 4-23. Means (Standard Deviations) of Performance Measures, Mathematical Processing - Baseline Trials.

| Level | Low | | | Medium | | | High | | |
|---|---|---|---|---|---|---|---|---|---|
| Trial | 06 | 08 | Both | 06 | 08 | Both | 06 | 08 | Both |
| MNO | 564<br>(179) | 540<br>(190) | 552<br>(185) | 1522<br>(567) | 1470<br>(592) | 1496<br>(579) | 2631<br>(990) | 2528<br>(997) | 2579<br>(993) |
| SDO | 279<br>(201) | 288<br>(285) | 284<br>(246) | 801<br>(431) | 824<br>(535) | 813<br>(485) | 1229<br>(540) | 1257<br>(625) | 1243<br>(583) |
| PCO | 97<br>(3) | 97<br>(2) | 97<br>(3) | 97<br>(3) | 97<br>(3) | 97<br>(3) | 97<br>(3) | 97<br>(5) | 97<br>(5) |
| STIMO | 157<br>(21) | 159<br>(24) | 158<br>(23) | 78<br>(17) | 80<br>(18) | 79<br>(17) | 51<br>(13) | 53<br>(13) | 52<br>(13) |
| MNP | 579<br>(193) | 565<br>(217) | 572<br>(205) | 1468<br>(582) | 1403<br>(566) | 1436<br>(574) | 2517<br>(1004) | 2365<br>(917) | 2441<br>(962) |
| SDP | 275<br>(197) | 299<br>(324) | 287<br>(268) | 773<br>(449) | 773<br>(508) | 773<br>(478) | 1171<br>(547) | 1121<br>(529) | 1146<br>(538) |
| PCP | 97<br>(3) | 97<br>(2) | 97<br>(3) | 97<br>(3) | 97<br>(3) | 97<br>(3) | 98<br>(3) | 97<br>(8) | 98<br>(6) |
| MNN | 550<br>(175) | 518<br>(169) | 534<br>(172) | 1571<br>(578) | 1530<br>(651) | 1551<br>(614) | 2789<br>(1038) | 2744<br>(1170) | 2767<br>(1104) |
| SDN | 269<br>(212) | 262<br>(249) | 265<br>(231) | 792<br>(459) | 834<br>(596) | 813<br>(531) | 1235<br>(633) | 1312<br>(769) | 1273<br>(704) |
| PCN | 97<br>(3) | 96<br>(3) | 96<br>(3) | 97<br>(3) | 97<br>(4) | 97<br>(4) | 96<br>(6) | 95<br>(6) | 96<br>(6) |

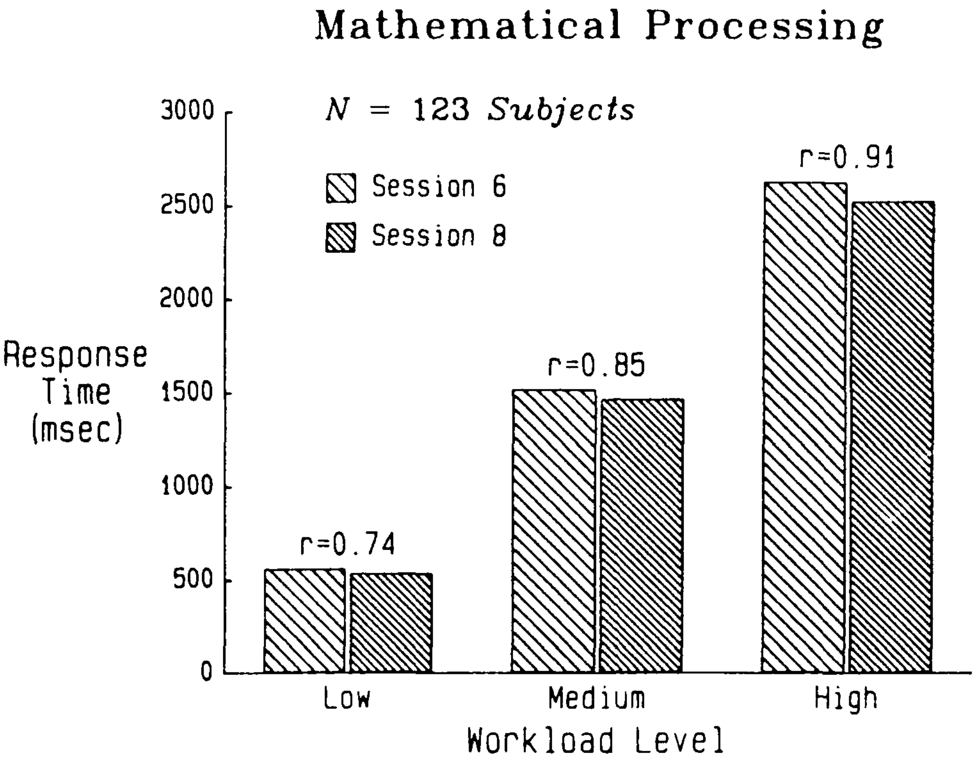


Figure 4-25. Mean Response Time for Mathematical Processing - Trials 6 and 8.

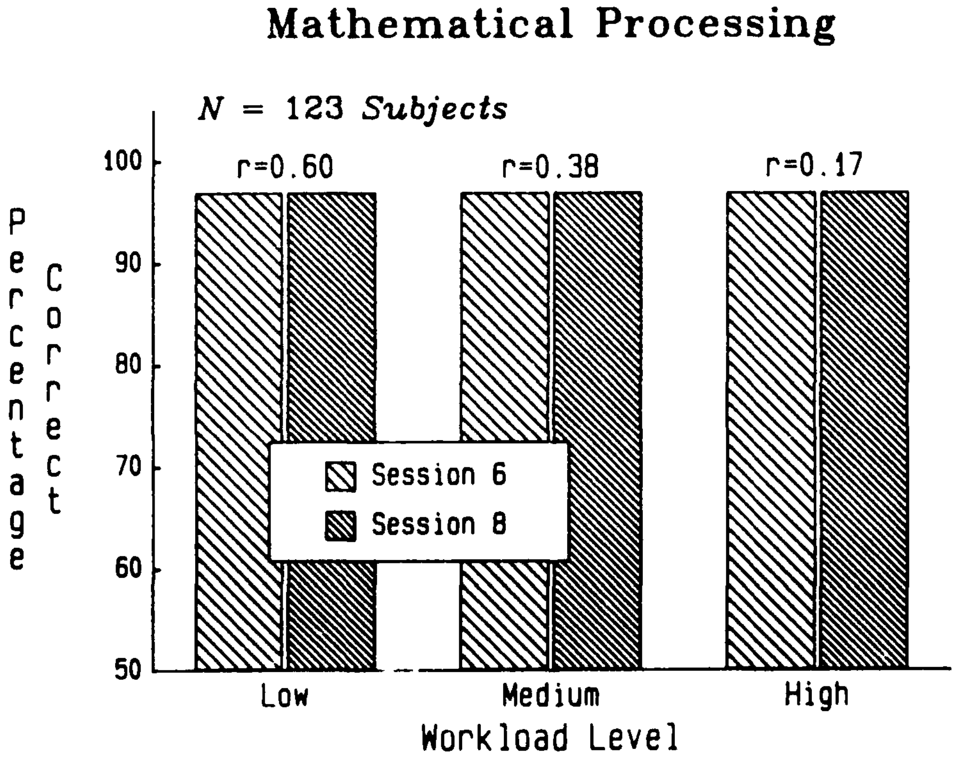


Figure 4-26. Percentage Correct for Mathematical Processing - Trials 6 and 8.

Table 4-24. ANOVA Summary for Level and Trial Effects, Mathematical Processing - Baseline Trials.

| Response Time | | | | | |
|---|---|---|---|---|---|
| Source | DF | Sum of Squares (x1000) | Mean Square (x1000) | F | p > F |
| Level (L) | 2 | 506171 | 253086 | 645.51 | 0.0001 |
| Trial (T) | 1 | 651 | 651 | 7.21 | 0.0083 |
| L by T | 2 | 201 | 100 | 3.45 | 0.0335 |
| Subject (S) | 122 | 217457 | 1782 | 61.14 | 0.0001 |
| L by S | 244 | 95666 | 392 | 13.45 | 0.0001 |
| T by S | 122 | 11009 | 90 | 3.10 | 0.0001 |
| Error | 244 | 7114 | 29 | | |
| Proportion Correct | | | | | |
| Source | DF | Sum of Squares (x0.001) | Mean Square (x0.001) | F | p > F |
| Level (L) | 2 | 5 | 3 | 3.41 | 0.0347 |
| Trial (T) | 1 | 2 | 2 | 1.58 | 0.2115 |
| L by T | 2 | 0 | 0 | 0.14 | 0.8665 |
| Subject (S) | 122 | 367 | 3 | 5.75 | 0.0001 |
| L by S | 244 | 181 | 1 | 1.42 | 0.0033 |
| T by S | 122 | 178 | 1 | 2.78 | 0.0001 |
| Error | 244 | 128 | 1 | | |

### 4.5.2 Gender and Prototype Analyses

As described in Section 4.1.2 for CR, further analyses of gender and prototype differences were performed. The performance measures are presented separately for men and women in Table 4-25 and Figures 4-27 and 4-28. Although there were no differences in response time between men and women, the women achieved slightly higher accuracy (1 to 2%, p = 0.009) than the men. There were no significant interactions involving gender for either response variable.

In contrast with CR, GR and LP, the Effort prototype group had slower response times than the other groups although the difference was not significant. There were no differences for the proportion correct measure. Refer to Section 5.2 for a discussion of the prototype grouping.

**Table 4-25. Means (Standard Deviations) of Performance Measures by Gender, Mathematical Processing - Baseline Trials.**

| Level | Low | | | Medium | | | High | | |
|---|---|---|---|---|---|---|---|---|---|
| Gender | Fem | Male | Both | Fem | Male | Both | Fem | Male | Both |
| MNO | 514<br>(137) | 563<br>(196) | 552<br>(185) | 1451<br>(481) | 1510<br>(605) | 1496<br>(579) | 2727<br>(805) | 2535<br>(1040) | 2579<br>(993) |
| SDO | 236<br>(117) | 298<br>(271) | 284<br>(246) | 788<br>(438) | 820<br>(499) | 813<br>(485) | 1329<br>(542) | 1217<br>(594) | 1243<br>(583) |
| PCO | 98<br>(2) | 97<br>(3) | 97<br>(3) | 98<br>(2) | 97<br>(3) | 97<br>(3) | 98<br>(3) | 96<br>(5) | 97<br>(5) |
| STIMO | 163<br>(18) | 157<br>(24) | 158<br>(23) | 80<br>(14) | 79<br>(18) | 79<br>(17) | 49<br>(10) | 53<br>(14) | 52<br>(13) |
| MNP | 536<br>(153) | 583<br>(217) | 572<br>(205) | 1397<br>(519) | 1447<br>(590) | 1436<br>(574) | 2533<br>(744) | 2414<br>(1018) | 2441<br>(962) |
| SDP | 233<br>(142) | 303<br>(293) | 287<br>(268) | 727<br>(425) | 786<br>(493) | 773<br>(478) | 1217<br>(495) | 1125<br>(549) | 1146<br>(538) |
| PCP | 98<br>(2) | 97<br>(3) | 97<br>(3) | 98<br>(2) | 97<br>(3) | 97<br>(3) | 98<br>(3) | 97<br>(6) | 98<br>(6) |
| MNN | 494<br>(127) | 546<br>(182) | 534<br>(172) | 1495<br>(501) | 1567<br>(644) | 1551<br>(614) | 3004<br>(968) | 2697<br>(1134) | 2767<br>(1104) |
| SDN | 223<br>(108) | 278<br>(255) | 265<br>(231) | 804<br>(491) | 816<br>(543) | 813<br>(531) | 1374<br>(677) | 1244<br>(711) | 1273<br>(704) |
| PCN | 98<br>(3) | 96<br>(3) | 96<br>(3) | 99<br>(2) | 97<br>(4) | 97<br>(4) | 97<br>(5) | 95<br>(6) | 96<br>(6) |

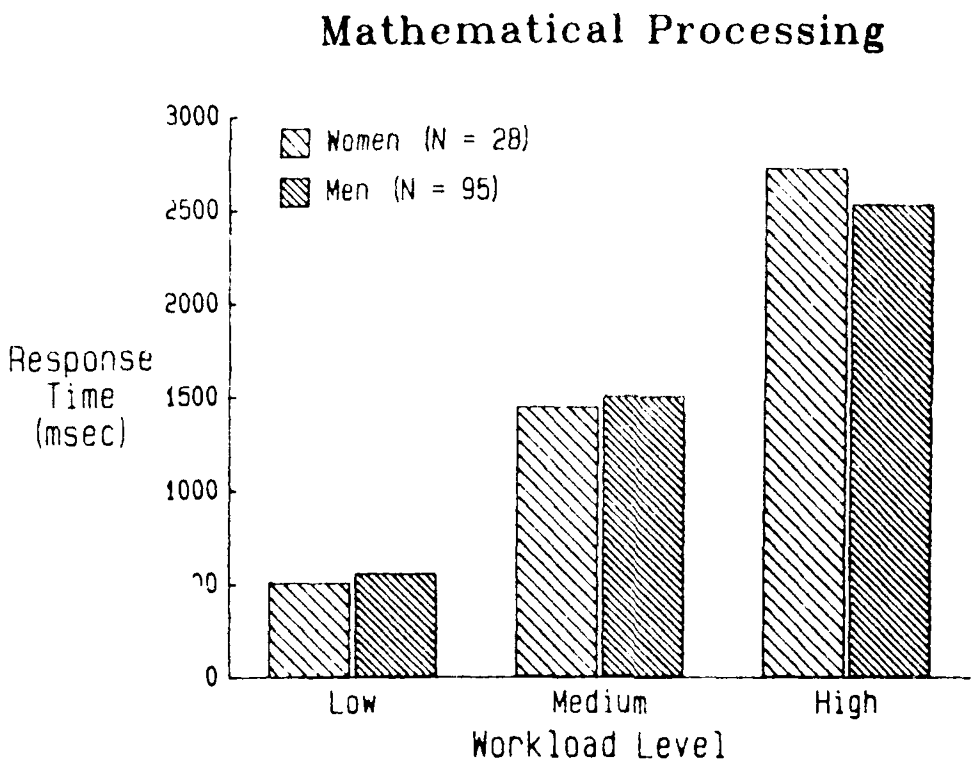


Figure 4-27. Mean Response Time for Mathematical Processing - Men vs. Women.

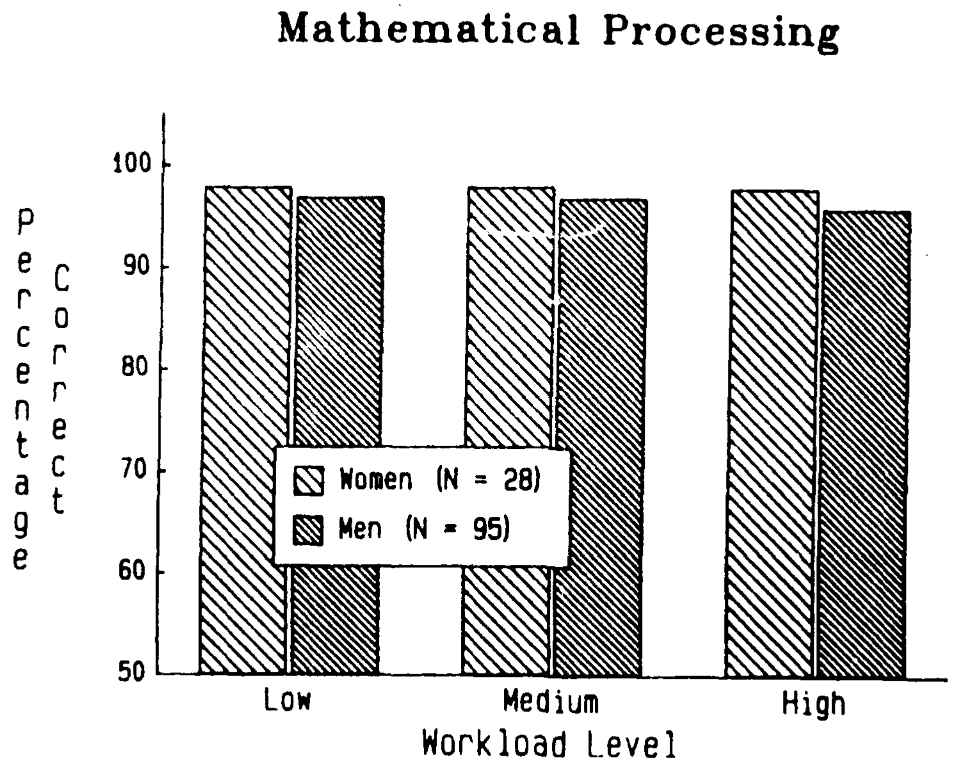


Figure 4-28. Percentage Correct for Mathematical Processing - Men vs. Women.

### 4.5.3 Training Data

The means and standard deviations of the major performance measures for Mathematical Processing for training Trials 1 through 5 are presented by difficulty level in Table 4-26. Response time and percentage correct are plotted in Figures 4-29 and 4-30. There was a steady improvement in speed but no practical change in accuracy during training for the MP task. For RT, there was a highly significant Trial by Level interaction with greater improvement at the higher difficulty levels.

Analysis of variance was used to determine significance between trials for RT and PC using the model presented in Section 4.1.1. Due to the significant Trial by Level interaction for RT ($p < 0.0001$), separate analyses were performed for each level using a reduced model involving only the trial and subject effects. A summary of the ANOVA results from the first set of analyses is presented in Table 4-27. The results of Tukey studentized range tests are summarized in Table 4-28. With respect to RT, there were no significant differences among Trials 3, 4 and 5 at the medium and high levels or between Trials 4 and 5 at the low level although the improvement trend continued. For PC, there were no differences among all five trials except at the medium level where Trial 1 demonstrated slightly lower accuracy than Trials 3 through 5.

## Table 4-26. Means (Standard Deviations) of Performance Measures, Mathematical Processing - Training Trials.

| Var. | Level | Trial | | | | |
|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 |
| MNO | Low | 897 ( 277) | 701 ( 198) | 643 ( 197) | 599 ( 195) | 589 ( 182) |
| | Med | 2184 ( 676) | 1895 ( 595) | 1721 ( 547) | 1633 ( 593) | 1616 ( 586) |
| | High | 3492 (1043) | 3091 (1056) | 2838 ( 987) | 2709 ( 996) | 2689 ( 939) |
| SDO | Low | 552 (284) | 368 (196) | 334 (222) | 311 (301) | 293 (171) |
| | Med | 1068 (431) | 942 (433) | 887 (397) | 844 (469) | 879 (449) |
| | High | 1513 (555) | 1337 (575) | 1232 (512) | 1200 (523) | 1285 (581) |
| PCO | Low | 96 (9) | 96 (12) | 97 (2) | 97 (2) | 97 (2) |
| | Med | 96 (5) | 96 (4) | 97 (3) | 97 (3) | 97 (3) |
| | High | 96 (5) | 97 (4) | 96 (4) | 96 (4) | 96 (3) |
| STIMO | Low | 124 (20) | 141 (20) | 148 (22) | 153 (21) | 154 (21) |
| | Med | 61 (12) | 67 (13) | 72 (14) | 75 (16) | 75 (17) |
| | High | 41 ( 9) | 45 (11) | 48 (11) | 50 (12) | 50 (12) |
| MNP | Low | 948 ( 323) | 721 ( 207) | 665 ( 219) | 621 ( 241) | 607 ( 199) |
| | Med | 2176 ( 726) | 1854 ( 615) | 1670 ( 558) | 1578 ( 624) | 1566 ( 634) |
| | High | 3398 (1039) | 2981 (1016) | 2719 ( 948) | 2560 ( 938) | 2554 ( 920) |
| MNN | Low | 851 ( 251) | 682 ( 200) | 622 ( 190) | 581 ( 179) | 574 ( 177) |
| | Med | 2208 ( 676) | 1941 ( 621) | 1779 ( 587) | 1689 ( 593) | 1665 ( 588) |
| | High | 3655 (1172) | 3249 (1194) | 3005 (1127) | 2906 (1118) | 2870 (1026) |

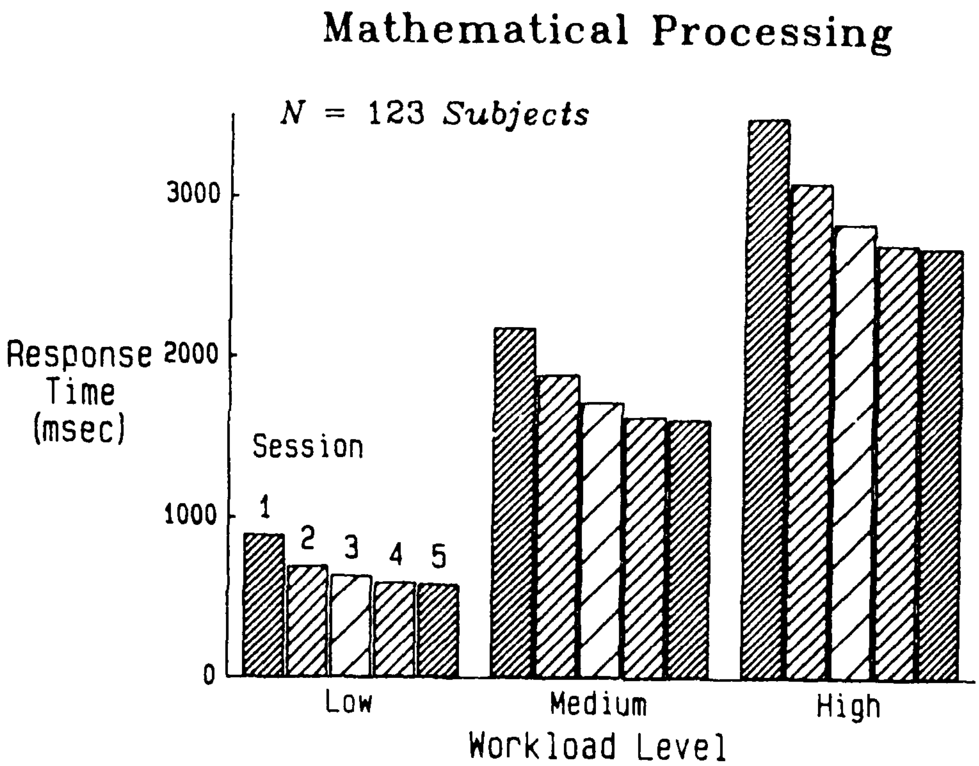


Figure 4-29. Mean Response Time for Mathematical Processing - Trials 1 through 5.

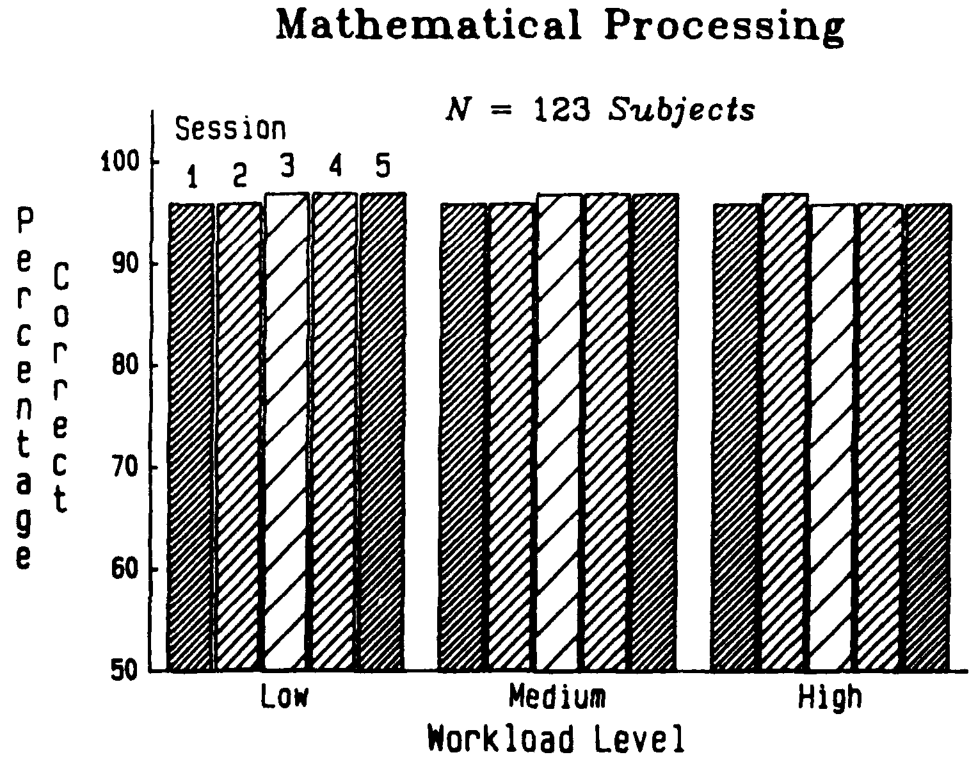


Figure 4-30. Percentage Correct for Mathematical Processing - Trials 1 through 5.

Table 4-27. ANOVA Summary for Level and Trial Effects, Mathematical Processing - Training Trials.

| Var. | Model | Level | | Trial | | Level * Trial | |
|---|---|---|---|---|---|---|---|
| | $R^2$ | $F_{(2,244)}$ | $p>F$ | $F_{(4,488)}$ | $p>F$ | $F_{(8,976)}$ | $p>F$ |
| MPMNO | 0.98 | 16389.52 | * | 414.27 | * | 26.22 | * |
| MPPCO | 0.58 | 2.36 | .0946 | 3.93 | .0036 | 0.92 | .4974 |

* $p < 0.0001$

Table 4-28. Significant ($\alpha = .01$) Trial Differences by Level.

| Var. | Level | $F_{(4,488)}$ | Trial |
|---|---|---|---|
| MNO | L | 181.17 | $1\ \ 2\ \ 3\ \ \underline{4\ \ 5}$ |
| | M | 96.18 | $1\ \ 2\ \ \underline{3\ \ 4\ \ 5}$ |
| | H | 94.89 | $1\ \ 2\ \ \underline{3\ \ 4\ \ 5}$ |
| PCO | L | 1.13 | $\underline{2\ \ 1\ \ 4\ \ 3\ \ 5}$ |
| | M | 5.36 | $\underline{1\ \ \underline{2}}\underline{\ \ 3\ \ 4\ \ 5}$ |
| | H | 2.38 | $\underline{1\ \ 3\ \ 5\ \ 4\ \ 2}$ |

## 4.6 Memory Search Task

The means and standard deviations for the Memory Search performance measures are presented in Table 4-29 for Trials 6 and 8. Overall mean response time (RT) and percentage correct (PC) are presented in Figures 4-31 and 4-32 with the intertrial correlations represented by the *r* values in the figures. Univariate summaries of the overall response time and proportion correct measures for the three difficulty levels are provided in Appendix A-6. Data for these summaries included the Trial 6 and Trial 8 data for all 123 subjects.

Response times were the fastest of all the central processing tasks at 0.4, 0.6 and 0.7 seconds for the three levels (positive set sizes 1, 4 and 6 respectively). The standard deviations increased from 0.07 to 0.16 seconds with increasing difficulty level. Proportion correct ranged from 0.97 to 0.89 with a similar increase in standard deviation from 0.03 to 0.07.

As with most of the central processing tasks, response time was faster (by 30 to 100 msec) for the *MATCH* (right button) responses compared with the *NON-MATCH* (left button) responses. In addition, proportion correct was noticeably higher for *MATCH* responses (0.93 vs. 0.85) at the high difficulty level but approximately the same for both buttons at the low and medium levels.

The *r* values were moderate for response time (0.59 to 0.75) but they were very low (0.05 to 0.47) for proportion correct. Again, this was probably due to the fast response times and high accuracy achieved by virtually all subjects at all levels.

### 4.6.1 Level and Trial Analyses

Analysis of variance was performed to verify the difficulty level manipulation and examine any trial differences using the model presented in Section 4.1.1. The results of the analyses for response time and proportion correct are summarized in Table 4-30. The model $R^2$ values were 0.95 for RT and 0.92 for PC. The Tukey studentized range test at $\alpha = 0.01$ demonstrated that all three difficulty levels differed significantly for both RT and PC. The mean response time for Trial 8 was marginally lower (p = 0.03) than the RT for Trial 6 with the largest difference occurring at the high level. The proportion correct values were essentially identical for Trials 6 and 8.

Table 4-29. Means (Standard Deviations) of Performance Measures, Memory Search - Baseline Trials.

| Level | Low | | | Medium | | | High | | |
|---|---|---|---|---|---|---|---|---|---|
| Trial | 06 | 08 | Both | 06 | 08 | Both | 06 | 08 | Both |
| MNO | 448<br>(65) | 442<br>(76) | 445<br>(71) | 601<br>(123) | 596<br>(137) | 598<br>(129) | 742<br>(173) | 709<br>(155) | 726<br>(164) |
| SDO | 143<br>(99) | 144<br>(102) | 143<br>(100) | 256<br>(173) | 249<br>(176) | 252<br>(174) | 439<br>(270) | 381<br>(226) | 410<br>(250) |
| PCO | 97<br>(4) | 97<br>(2) | 97<br>(3) | 96<br>(5) | 96<br>(4) | 96<br>(5) | 89<br>(7) | 90<br>(7) | 89<br>(7) |
| STIMO | 275<br>(29) | 279<br>(32) | 277<br>(31) | 225<br>(31) | 228<br>(35) | 226<br>(33) | 193<br>(33) | 200<br>(36) | 197<br>(34) |
| MNP | 436<br>(59) | 429<br>(71) | 433<br>(66) | 570<br>(106) | 560<br>(114) | 565<br>(110) | 688<br>(154) | 661<br>(139) | 674<br>(147) |
| SDP | 128<br>(81) | 125<br>(96) | 126<br>(89) | 214<br>(162) | 199<br>(149) | 207<br>(156) | 368<br>(273) | 321<br>(210) | 345<br>(244) |
| PCP | 97<br>(3) | 97<br>(3) | 97<br>(3) | 95<br>(4) | 95<br>(6) | 95<br>(5) | 94<br>(7) | 93<br>(8) | 93<br>(8) |
| MNN | 463<br>(98) | 457<br>(88) | 460<br>(93) | 641<br>(163) | 632<br>(164) | 636<br>(163) | 809<br>(219) | 767<br>(193) | 788<br>(207) |
| SDN | 147<br>(125) | 152<br>(115) | 150<br>(120) | 281<br>(201) | 279<br>(207) | 280<br>(204) | 474<br>(301) | 411<br>(273) | 442<br>(289) |
| PCN | 97<br>(8) | 98<br>(2) | 98<br>(6) | 96<br>(9) | 96<br>(4) | 96<br>(7) | 84<br>(13) | 86<br>(12) | 85<br>(12) |

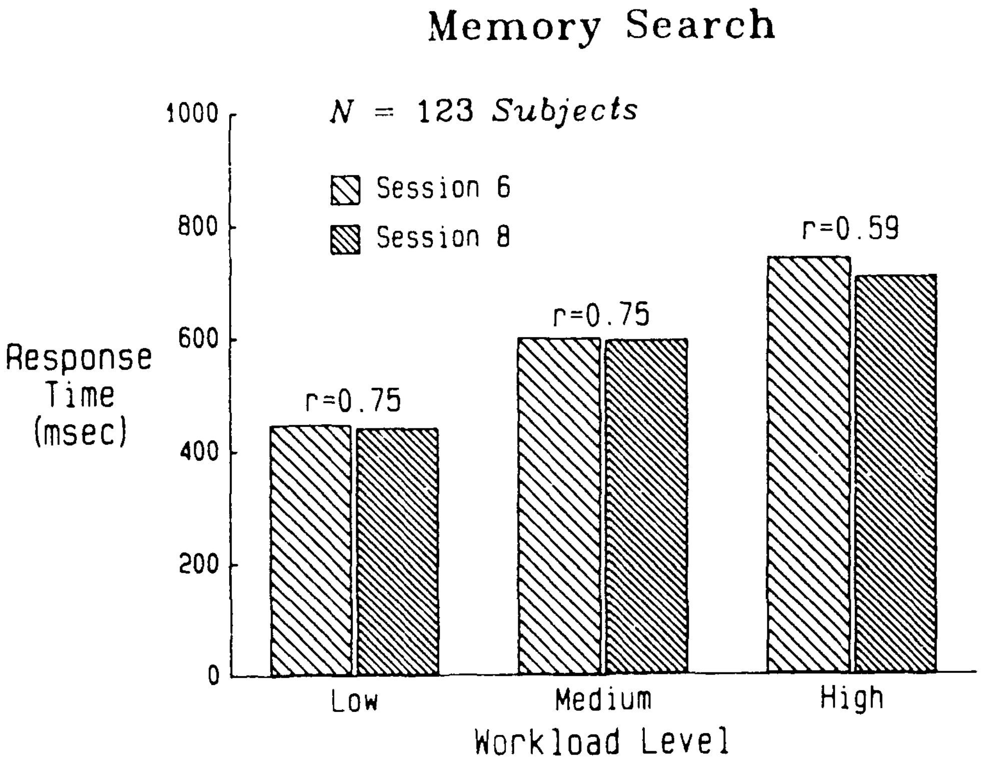


**Figure 4-31. Mean Response Time for Memory Search - Trials 6 and 8.**

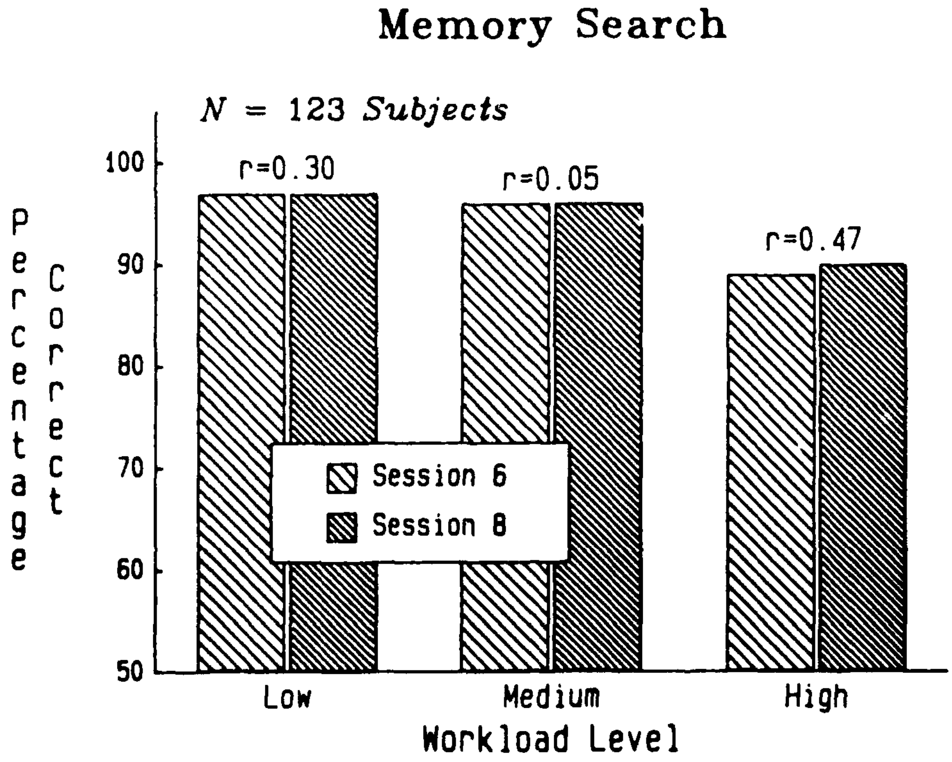


**Figure 4-32. Percentage Correct for Memory Search - Trials 6 and 8.**

Table 4-30. ANOVA Summary for Level and Trial Effects, Memory Search - Baseline Trials.

| Response Time | | | | | |
|---|---|---|---|---|---|
| Source | DF | Sum of Squares (x1000) | Mean Square (x1000) | F | p > F |
| Level (L) | 2 | 9733 | 4867 | 550.96 | 0.0001 |
| Trial (T) | 1 | 39 | 39 | 4.95 | 0.0280 |
| L by T | 2 | 31 | 18 | 3.52 | 0.0310 |
| Subject (S) | 122 | 7685 | 63 | 14.09 | 0.0001 |
| L by S | 244 | 2155 | 9 | 1.98 | 0.0001 |
| T by S | 122 | 959 | 8 | 1.76 | 0.0001 |
| Error | 244 | 1091 | 4 | | |

| Proportion Correct | | | | | |
|---|---|---|---|---|---|
| Source | DF | Sum of Squares (x0.001) | Mean Square (x0.001) | F | p > F |
| Level (L) | 2 | 868 | 434 | 189.92 | 0.0001 |
| Trial (T) | 1 | 0 | 0 | 0.12 | 0.7317 |
| L by T | 2 | 1 | 1 | 0.66 | 0.5910 |
| Subject (S) | 122 | 705 | 6 | 6.05 | 0.0001 |
| L by S | 244 | 558 | 2 | 2.39 | 0.0001 |
| T by S | 122 | 400 | 3 | 3.43 | 0.0001 |
| Error | 244 | 233 | 1 | | |

### 4.6.2 Gender and Prototype Analyses

As described in Section 4.1.2 for CR, further analyses of gender and prototype differences were performed. The performance measures are presented separately for men and women in Table 4-31 and Figures 4-33 and 4-34. Although not significant across all three levels, women responded faster than men especially at the high level. However, the accuracy for women was also slightly lower than for men. The Level by Gender and Trial by Gender interactions were significant for both response variables.

With respect to prototype, there were no differences or significant interactions for either variable. Refer to Section 5.2 for a discussion of the prototype grouping.

Table 4-31. Means (Standard Deviations) of Performance Measures by Gender, Memory Search - Baseline Trials.

| Level | Low | | | Medium | | | High | | |
|---|---|---|---|---|---|---|---|---|---|
| Gender | Fem | Male | Both | Fem | Male | Both | Fem | Male | Both |
| MNO | 447<br>(85) | 444<br>(66) | 445<br>(71) | 575<br>(106) | 605<br>(135) | 598<br>(129) | 670<br>(147) | 742<br>(166) | 726<br>(164) |
| SDO | 131<br>(68) | 147<br>(107) | 143<br>(100) | 230<br>(136) | 259<br>(183) | 252<br>(174) | 379<br>(250) | 420<br>(250) | 410<br>(250) |
| PCO | 98<br>(5) | 97<br>(2) | 97<br>(3) | 96<br>(6) | 96<br>(4) | 96<br>(5) | 86<br>(7) | 90<br>(7) | 89<br>(7) |
| STIMO | 271<br>(33) | 279<br>(30) | 277<br>(31) | 227<br>(29) | 226<br>(34) | 226<br>(33) | 205<br>(32) | 194<br>(35) | 197<br>(34) |
| MNP | 438<br>(77) | 431<br>(62) | 433<br>(66) | 546<br>(95) | 571<br>(114) | 565<br>(110) | 629<br>(141) | 688<br>(146) | 674<br>(147) |
| SDP | 118<br>(58) | 128<br>(96) | 126<br>(89) | 193<br>(118) | 211<br>(165) | 207<br>(156) | 332<br>(280) | 348<br>(233) | 345<br>(244) |
| PCP | 98<br>(3) | 97<br>(3) | 97<br>(3) | 97<br>(3) | 95<br>(6) | 95<br>(5) | 95<br>(8) | 93<br>(8) | 93<br>(8) |
| MNN | 466<br>(131) | 459<br>(78) | 460<br>(93) | 621<br>(167) | 641<br>(162) | 636<br>(163) | 735<br>(188) | 804<br>(210) | 788<br>(207) |
| SDN | 137<br>(86) | 153<br>(128) | 150<br>(120) | 255<br>(193) | 287<br>(206) | 280<br>(204) | 400<br>(263) | 455<br>(295) | 442<br>(289) |
| PCN | 97<br>(11) | 98<br>(2) | 98<br>(6) | 94<br>(12) | 96<br>(4) | 96<br>(7) | 78<br>(12) | 87<br>(12) | 85<br>(12) |

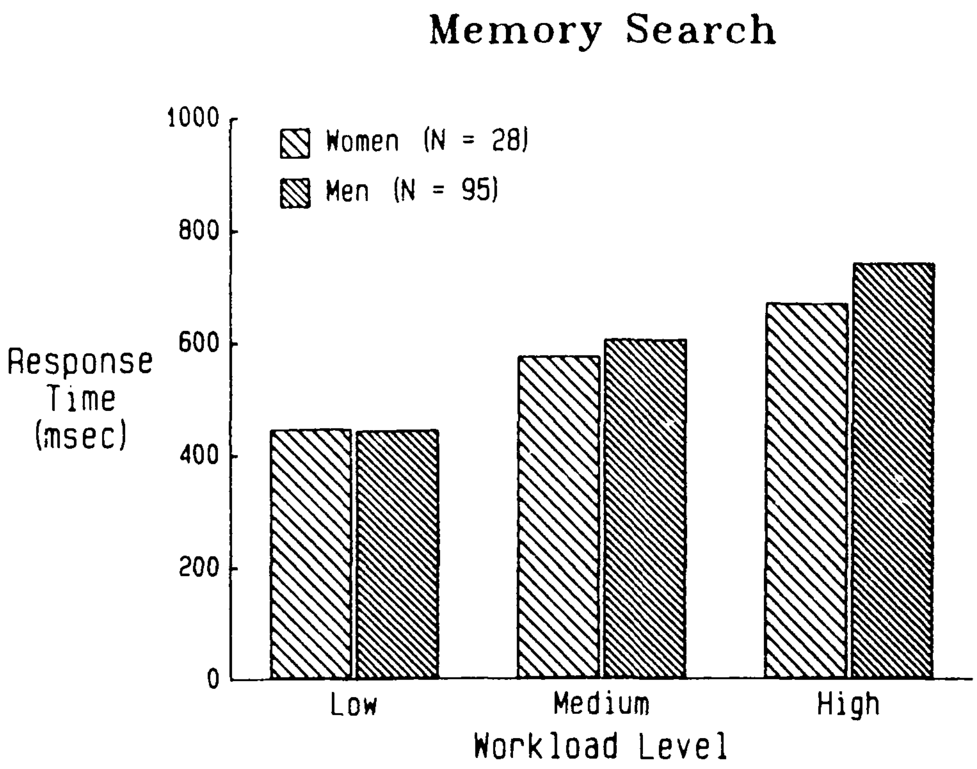


Figure 4-33. Mean Response Time for Memory Search - Men vs. Women.

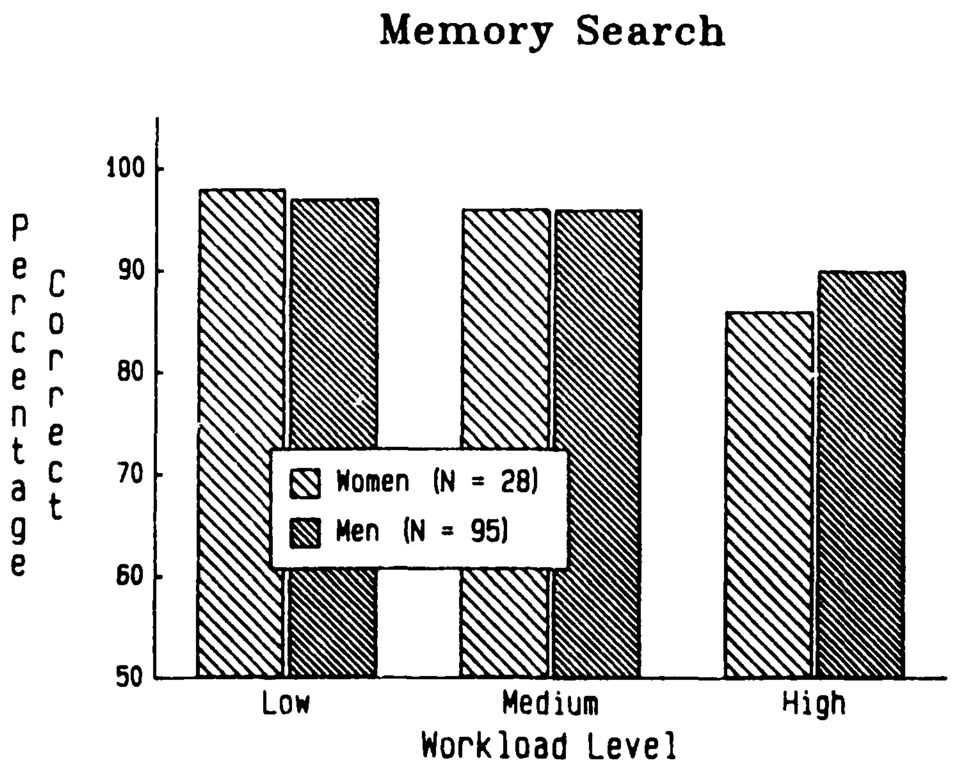


Figure 4-34. Percentage Correct for Memory Search - Men vs. Women.

### 4.6.3 Training Data

The means and standard deviations of the major performance measures for Memory Search for training Trials 1 through 5 are presented by difficulty level in Table 4-32. Response time and percentage correct are plotted in Figures 4-35 and 4-36. As with MP, there was a steady improvement in speed but no practical change in accuracy during training for the MS task.

Analysis of variance was used to determine significance between trials for RT and PC using the model presented in Section 4.1.1. Due to the significant (p = 0.015) Trial by Level interaction for PC, separate analyses were performed for each level using a reduced model involving only the trial and subject effects. A summary of the ANOVA results from the first set of analyses is presented in Table 4-33. The results of Tukey studentized range tests are summarized in Table 4-34. For RT, there was no significant difference between Trials 4 and 5 at any level although times continued to show improvement. For PC, there were no differences among all five trials at any level.

Table 4-32. Means (Standard Deviations) of Performance Measures, Memory Search - Training Trials.

| Var. | Level | Trial | | | | |
|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 |
| MNO | Low | 561 (105) | 502 ( 79) | 479 ( 76) | 461 ( 78) | 455 ( 72) |
| | Med | 712 (131) | 655 (110) | 637 (118) | 612 (116) | 599 (101) |
| | High | 835 (192) | 785 (169) | 764 (159) | 741 (168) | 724 (151) |
| SDO | Low | 219 (134) | 163 ( 90) | 149 ( 86) | 147 ( 91) | 151 (119) |
| | Med | 308 (129) | 278 (144) | 268 (166) | 243 (134) | 240 (136) |
| | High | 452 (224) | 443 (266) | 419 (237) | 412 (245) | 398 (238) |
| PCO | Low | 96 (9) | 97 (4) | 98 (4) | 97 (5) | 97 (4) |
| | Med | 96 (5) | 96 (4) | 95 (5) | 96 (4) | 96 (4) |
| | High | 88 (8) | 89 (8) | 89 (8) | 88 (8) | 89 (7) |
| STIMO | Low | 230 (30) | 251 (30) | 260 (29) | 269 (32) | 271 (29) |
| | Med | 195 (28) | 208 (27) | 214 (29) | 221 (30) | 224 (29) |
| | High | 175 (29) | 183 (29) | 187 (30) | 193 (32) | 195 (32) |
| MNP | Low | 536 ( 95) | 485 ( 67) | 466 ( 64) | 447 ( 67) | 442 ( 64) |
| | Med | 674 (109) | 617 ( 91) | 599 (107) | 581 (107) | 566 ( 84) |
| | High | 784 (182) | 739 (154) | 715 (148) | 691 (140) | 675 (131) |
| MNN | Low | 587 (149) | 524 (124) | 496 (110) | 479 (106) | 471 (100) |
| | Med | 753 (163) | 700 (154) | 678 (148) | 647 (140) | 636 (140) |
| | High | 894 (224) | 843 (212) | 828 (196) | 803 (218) | 787 (192) |

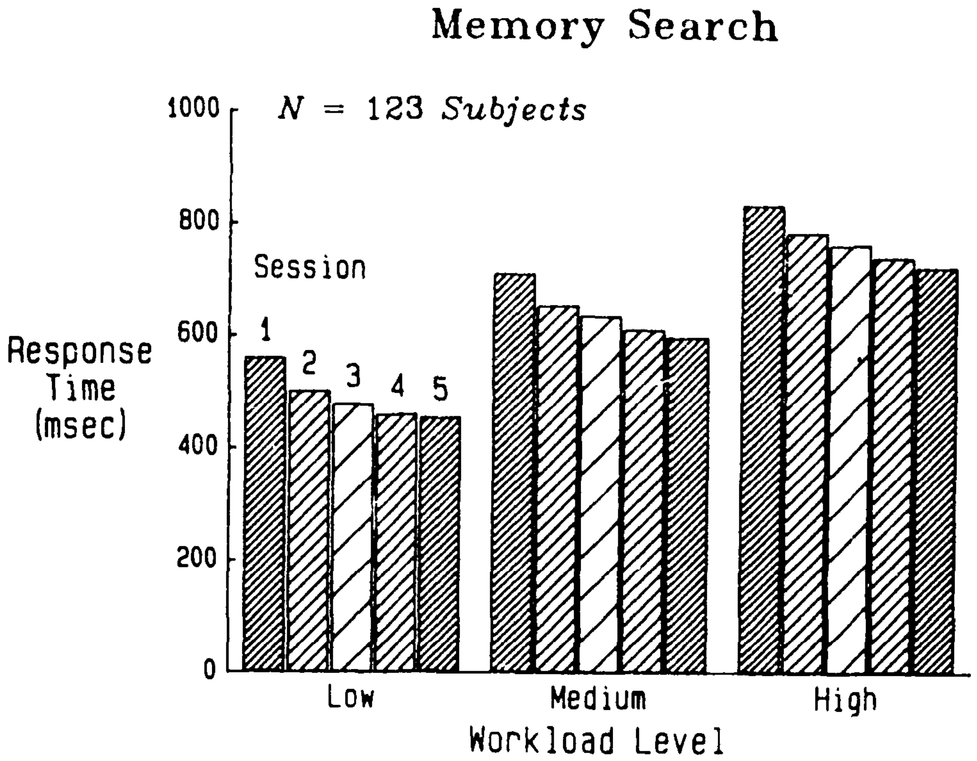


Figure 4-35. Mean Response Time for Memory Search - Trials 1 through 5.

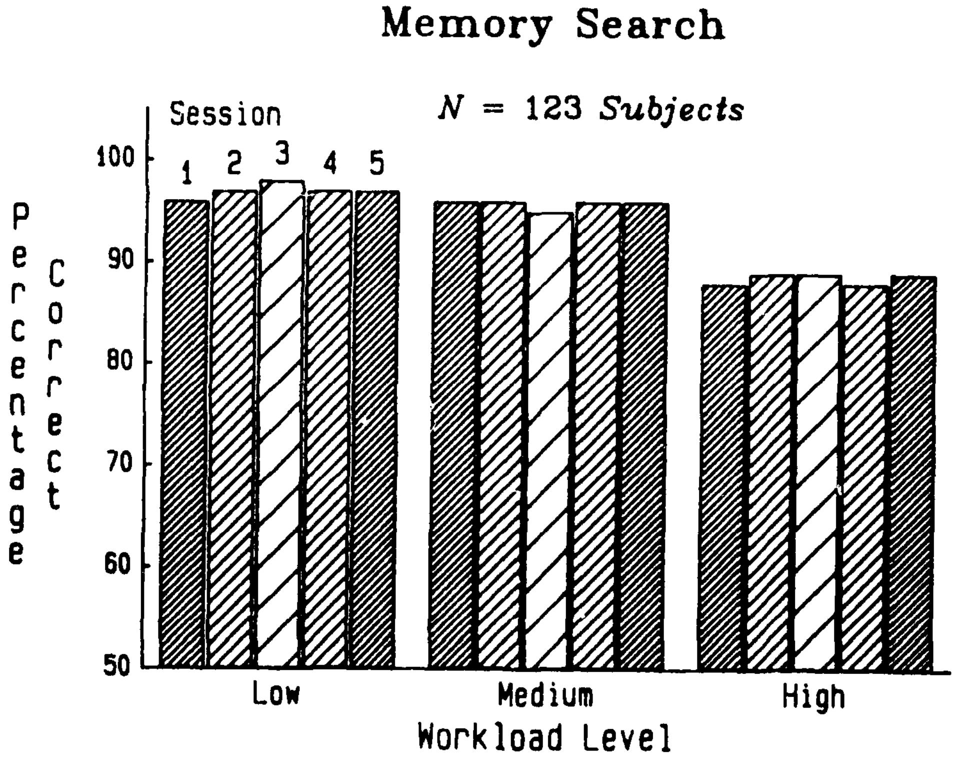


Figure 4-36. Percentage Correct for Memory Search - Trials 1 through 5.

Table 4-33. ANOVA Summary for Level and Trial Effects, Memory Search - Training Trials.

| Var. | Model | Level | | Trial | | Level * Trial | |
|---|---|---|---|---|---|---|---|
| | $R^2$ | $F_{(2,244)}$ | $p>F$ | $F_{(4,488)}$ | $p>F$ | $F_{(8,976)}$ | $p>F$ |
| MSMNO | 0.93 | 717.13 | * | 82.74 | * | 0.37 | .9363 |
| MSPCO | 0.82 | 257.36 | * | 1.68 | .1527 | 2.38 | .0153 |

* p < 0.0001

Table 4-34. Significant (α = .01) Trial Differences by Level.

| Var. | Level | $F_{(4,488)}$ | Trial | | | | |
|---|---|---|---|---|---|---|---|
| | L | 92.80 | 1 | 2 | 3 | 4 | 5 |
| MNO | M | 59.58 | 1 | 2 | 3 | 4 | 5 |
| | H | 22.24 | 1 | 2 | 3 | 4 | 5 |
| | L | 2.84 | 1 | 5 | 4 | 2 | 3 |
| PCO | M | 1.15 | 3 | 5 | 4 | 1 | 2 |
| | H | 1.94 | 4 | 1 | 3 | 2 | 5 |

## 4.7 Probability Monitoring Task

The Probability Monitoring task is the major CTS task that emphasizes the input stage of information processing. Subjects monitor one, three or four dials looking for a signal bias during which the pointer on one of the dials spends a higher percentage of time (95%, 85% or 75%) on one side of the dial center line. One drawback of the first version of the CTS PM task is the low number of signals (2 to 3) that occur during a three-minute trial. Thus, average response times for the trial and the proportion of correct detections are based on an extremely small or zero (with no detections) sample size.

To overcome this difficulty in the current study, data was collapsed across subjects or across trials by computing the mean response time for all signals in a given category (trial, gender, prototype, etc.). Percentage correct was computed in a similar fashion by dividing the total number of correct detections by the total number of signals in a given category.

The means for the Probability Monitoring performance measures for Trials 6 and 8 are presented in Table 4-35. No standard deviations or intertrial correlations are presented due to the method of collapsing the data as mentioned above. Mean response time (PMRT) and the percentage of correct detections (PMPC) are plotted in Figures 4-37 and 4-38 respectively. The average false alarm rate (PMFA) and average number of signals per trial (PMTS) are presented in Figures 4-39 and 4-40. Univariate summaries of the response time, proportion of correct detections and false alarms for the three difficulty levels are provided in Appendix A-7. The first set of summaries is based on the raw Trial 6 and Trial 8 data for all 123 subjects. A second set of summaries was produced by first collapsing the data for each subject across Trials 5, 6 and 8 (the last training day plus the two baseline days), and computing average RT, PC and FA values as described previously.

Response times were approximately 8, 16 and 18 seconds for the low, medium and high levels respectively. The larger gap between the low and medium levels indicated a greater change in task difficulty between these levels. The proportion of signal detections was very high (0.97) at the low difficulty level but dropped off sharply at the medium (0.79) and high (0.42) levels. The number of false alarms increased from an average of 0.3 per trial at the low level to 0.9 per trial at the medium level and 1.6 per trial at the high level.

An unfortunate bias in the CTS software was discovered in that the number of total signals per trial was not constant at all three difficulty levels. Rather, it dropped from a mean of 2.9 signals per trial at the low level to 2.7 at the medium level and 2.2 at the high level. At the low level, almost all trials contained three signals. At the high level, most trials contained only two signals, adding to the problem of obtaining meaningful data on an individual trial basis.

**Table 4-35. Means of Performance Measures, Probability Monitoring - Baseline Trials.**

| Level | Low | | | Medium | | | High | | |
|---|---|---|---|---|---|---|---|---|---|
| **Trial** | **06** | **08** | **Both** | **06** | **08** | **Both** | **06** | **08** | **Both** |
| PMRT | 8.5 | 8.2 | 8.4 | 15.9 | 16.0 | 15.9 | 18.3 | 16.7 | 17.5 |
| PMPC | 0.98 | 0.98 | 0.98 | 0.79 | 0.79 | 0.79 | 0.43 | 0.40 | 0.42 |
| PMFA | 0.41 | 0.24 | 0.32 | 0.86 | 0.89 | 0.88 | 1.54 | 1.66 | 1.60 |
| PMTS | 2.89 | 2.91 | 2.90 | 2.72 | 2.69 | 2.70 | 2.11 | 2.19 | 2.15 |

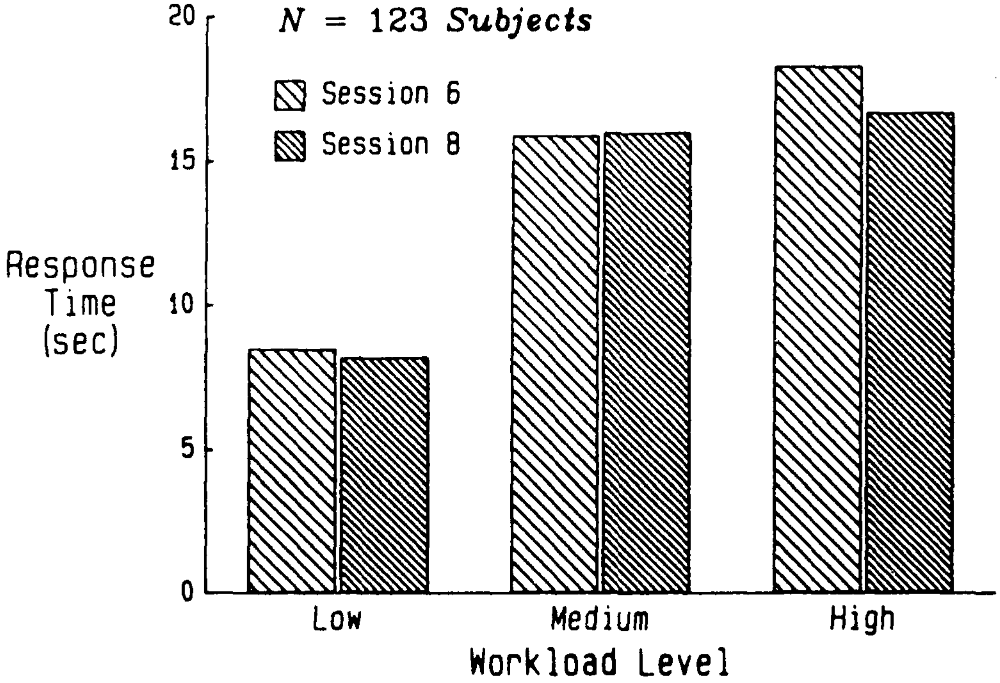


**Figure 4-37. Response Time for Probability Monitoring - Baseline Trials.**

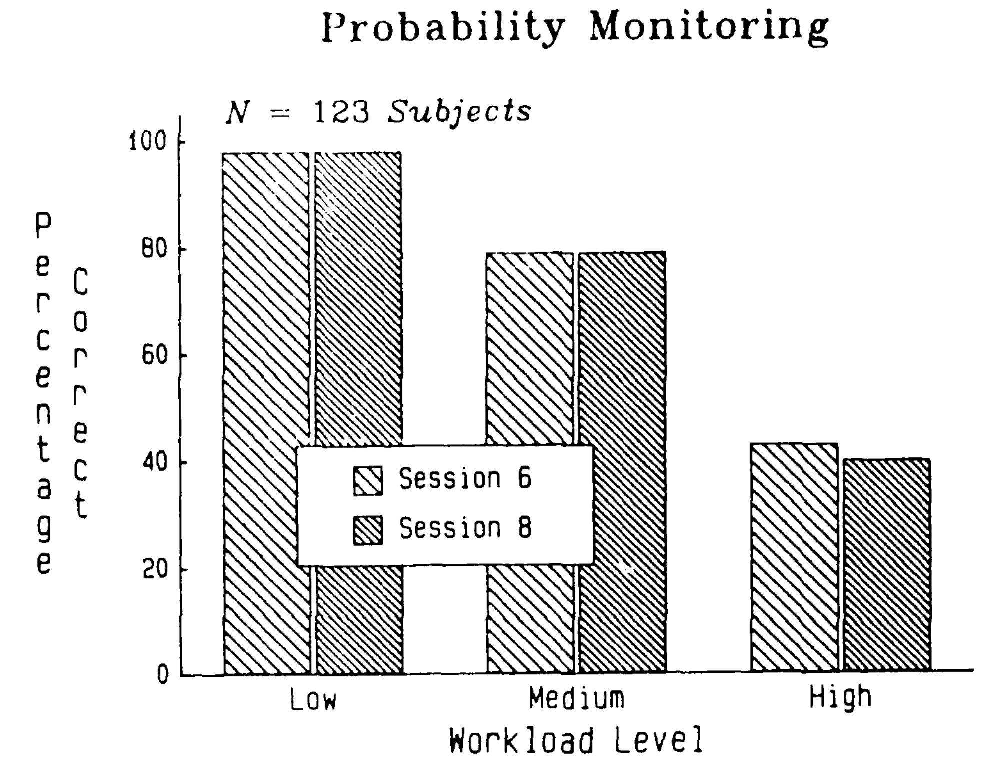


Figure 4-38. Percent Correct Detections for Probability Monitoring - Baseline Trials.

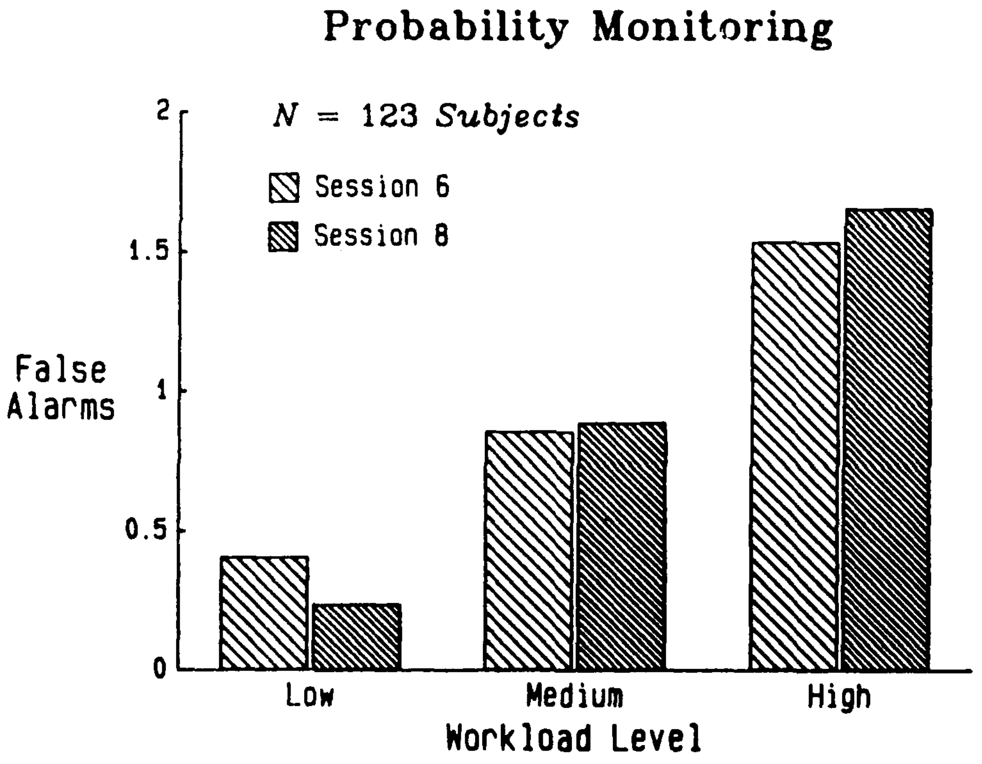


Figure 4-39. False Alarms per Trial for Probability Monitoring - Baseline Trials.

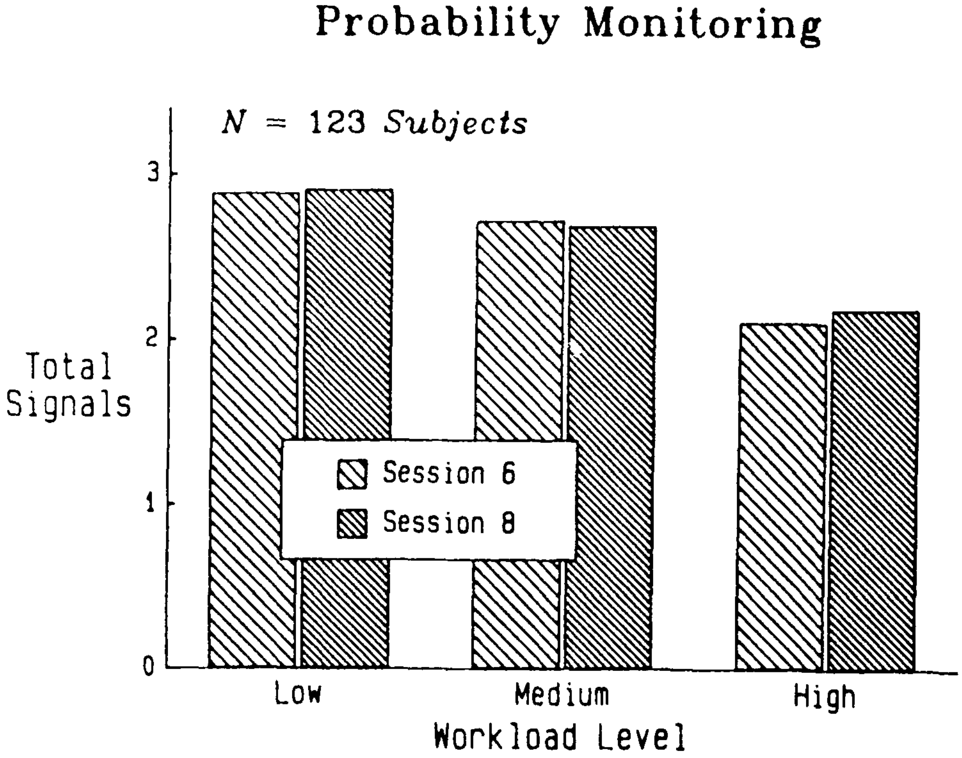


**Figure 4-40. Total Signals per Trial for Probability Monitoring - Baseline Trials.**

### 4.7.1 Level and Trial Analyses

Two separate analyses of variance were performed to verify the difficulty level manipulation and examine any trial differences. The first analysis used the model presented in Section 4.1.1 and the raw, uncollapsed data. The results of this analysis are presented in Table 4-36. The model $R^2$ values were 0.90 for PMRT, 0.84 for PMPC and 0.87 for PMFA. The Tukey studentized range test at $\alpha = 0.01$ indicated a significant difference in PMRT between the low and both medium and high levels, but there was no significant difference in PMRT between the medium and high levels. However, all three difficulty levels differed significantly for PMPC, PMFA and PMTS. There were no differences between Trials 6 and 8 for any of the variables that were examined.

Since there were no significant differences between trials, a second set of analyses were performed with data collapsed across Trials 5, 6 and 8 for each subject. The ANOVA model was an additive model involving Level and Subject as factors. The results of the analyses (Table 4-37) and of the Tukey tests did not differ from those of the first analysis.

**Table 4-36. ANOVA Summary for Level and Trial Effects, Probability Monitoring - Baseline Trials.**

Response Time

| Source | DF | Sum of Squares | Mean Square | F | p > F |
|---|---|---|---|---|---|
| Level (L) | 2 | 10976 | 5488 | 145.07 | 0.0001 |
| Trial (T) | 1 | 46 | 46 | 1.31 | 0.2554 |
| L by T | 2 | 138 | 69 | 3.60 | 0.0296 |
| Subject (S) | 122 | 6243 | 51 | 2.66 | 0.0001 |
| L by S | 232 | 8777 | 38 | 1.97 | 0.0001 |
| T by S | 122 | 4316 | 35 | 1.84 | 0.0001 |
| Error | 167 | 3210 | 19 | | |

Proportion Correct

| Source | DF | Sum of Squares (x0.01) | Mean Square (x0.01) | F | p > F |
|---|---|---|---|---|---|
| Level (L) | 2 | 4274 | 2137 | 361.95 | 0.0001 |
| Trial (T) | 1 | 0 | 0 | 0.07 | 0.7850 |
| L by T | 2 | 6 | 3 | 0.49 | 0.6154 |
| Subject (S) | 122 | 1029 | 8 | 1.43 | 0.0098 |
| L by S | 244 | 1441 | 6 | 1.00 | 0.4997 |
| T by S | 122 | 614 | 5 | 0.85 | 0.8378 |
| Error | 244 | 1441 | 6 | | |

False Alarms

| Source | DF | Sum of Squares | Mean Square | F | p > F |
|---|---|---|---|---|---|
| Level (L) | 2 | 203 | 101 | 101.06 | 0.0001 |
| Trial (T) | 1 | 0 | 0 | 0.02 | 0.9015 |
| L by T | 2 | 3 | 1 | 1.93 | 0.1474 |
| Subject (S) | 122 | 530 | 4 | 6.34 | 0.0001 |
| L by S | 244 | 245 | 1 | 1.47 | 0.0015 |
| T by S | 122 | 97 | 1 | 1.16 | 0.1669 |
| Error | 244 | 167 | 1 | | |

Table 4-36. ANOVA Summary for Level and Trial Effects, Probability Monitoring - Baseline Trials (continued).

| Total Signals | | | | | |
|---|---|---|---|---|---|
| Source | DF | Sum of Squares (x0.1) | Mean Square (x0.1) | F | p > F |
| Level (L) | 2 | 741 | 370 | 263.62 | 0.0001 |
| Trial (T) | 1 | 1 | 1 | 0.70 | 0.4035 |
| L by T | 2 | 3 | 1 | 1.11 | 0.3298 |
| Subject (S) | 122 | 195 | 2 | 1.21 | 0.1024 |
| L by S | 244 | 343 | 1 | 1.07 | 0.2996 |
| T by S | 122 | 191 | 2 | 1.19 | 0.1286 |
| Error | 244 | 320 | 1 | | |

Table 4-37. ANOVA Summary for Level Effect, Probability Monitoring - Collapsed Data.

| Var. | Model $R^2$ | Sum of Squares | Mean Square | $F_{(2,244)}$ | p > F |
|---|---|---|---|---|---|
| PMRT | 0.65 | 5641.54 | 2820.77 | 140.69 | * |
| PMPC | 0.85 | 21.56 | 10.78 | 587.69 | * |
| PMFA | 0.80 | 113.23 | 56.61 | 138.29 | * |
| PMTS | 0.79 | 35.96 | 17.98 | 380.81 | * |

* p < 0.0001

### 4.7.2 Gender and Prototype Analyses

For Trials 6 and 8, a data set was created by collapsing across all female subjects and all male subjects for each trial. For each of the response variables, this provided a mean value for Trial 6 - Females, Trial 6 - Males, Trial 8 - Females and Trial 8 - Males. The means for females and males are presented in Table 4-38 and Figures 4-41 through 4-43. Analysis of variance was performed using a different data set based on collapsing data for individual subjects across Trials 5, 6 and 8. Women produced a significantly higher (p = 0.04) rate of false alarms than men (1.3 vs. 0.9) along with a significant (p = 0.03) Level by Gender interaction. There were no other significant differences among the response variables.

With respect to prototype, there were no differences or significant interactions for any variable. Refer to Section 5.2 for a discussion of the prototype grouping.

**Table 4-38. Means of Performance Measures by Gender, Probability Monitoring - Baseline Trials.**

| Level | Low | | Medium | | High | |
|---|---|---|---|---|---|---|
| Gender | Fem | Male | Fem | Male | Fem | Male |
| PMRT | 8.6 | 8.3 | 15.5 | 16.1 | 16.4 | 17.9 |
| PMPC | 0.99 | 0.98 | 0.79 | 0.79 | 0.45 | 0.41 |
| PMFA | 0.39 | 0.30 | 1.20 | 0.78 | 1.87 | 1.52 |
| PMTS | 2.89 | 2.90 | 2.71 | 2.70 | 2.16 | 2.15 |

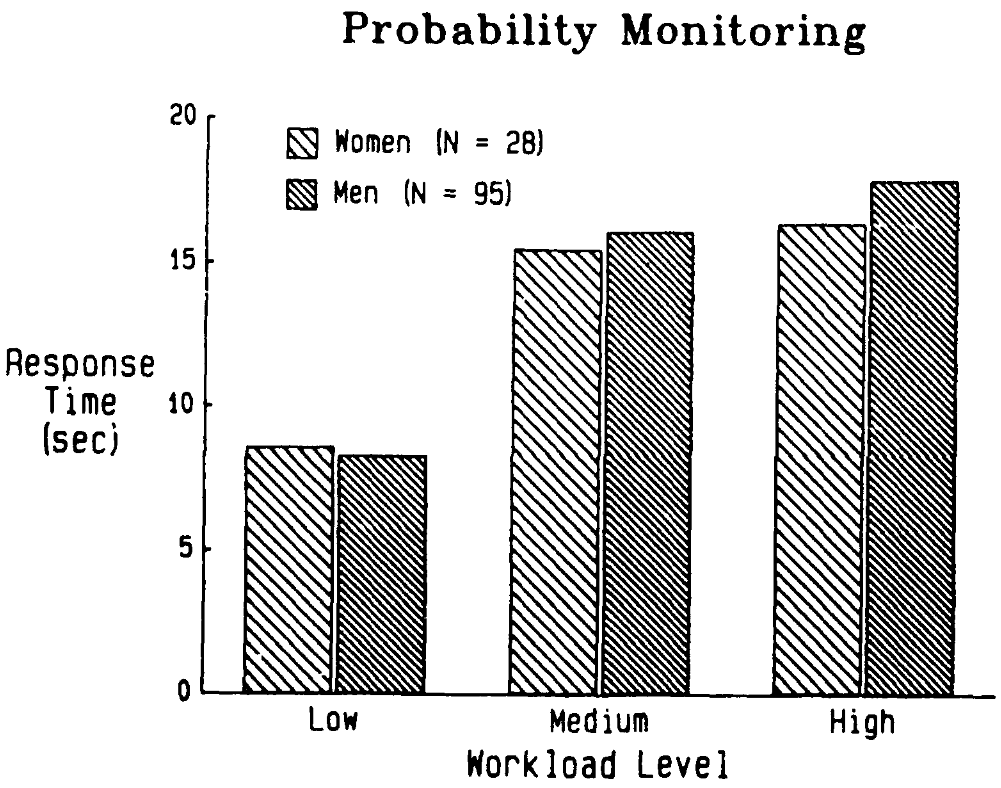


**Figure 4-41. Response Time for Probability Monitoring - Men vs. Women.**

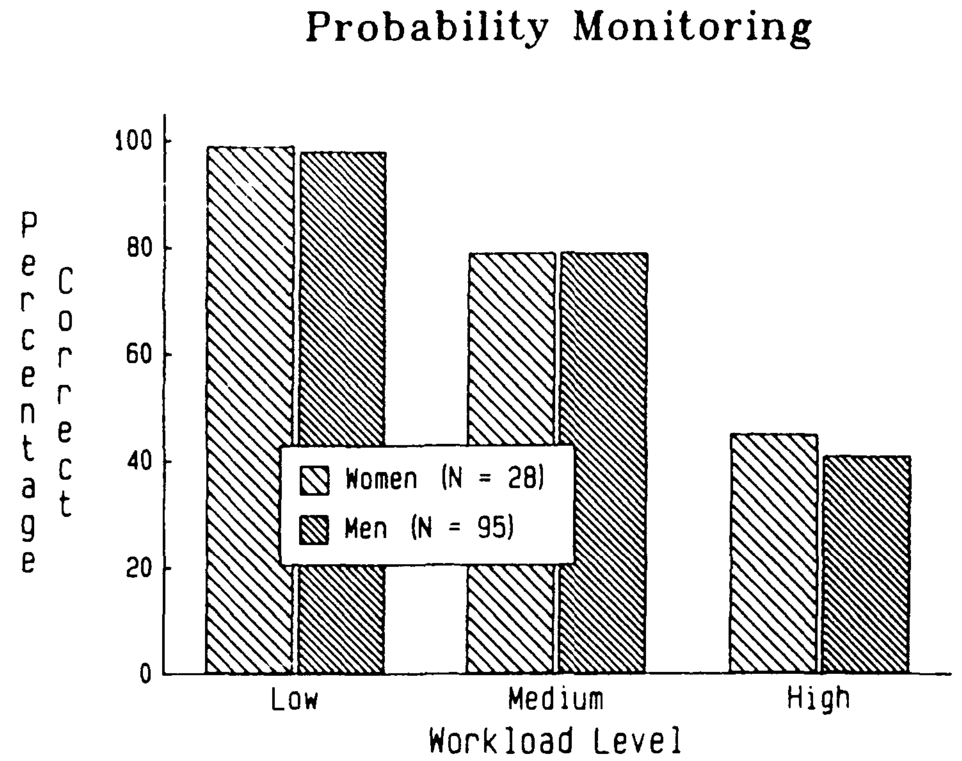


Figure 4-42. Percent Correct for Probability Monitoring - Men vs. Women.

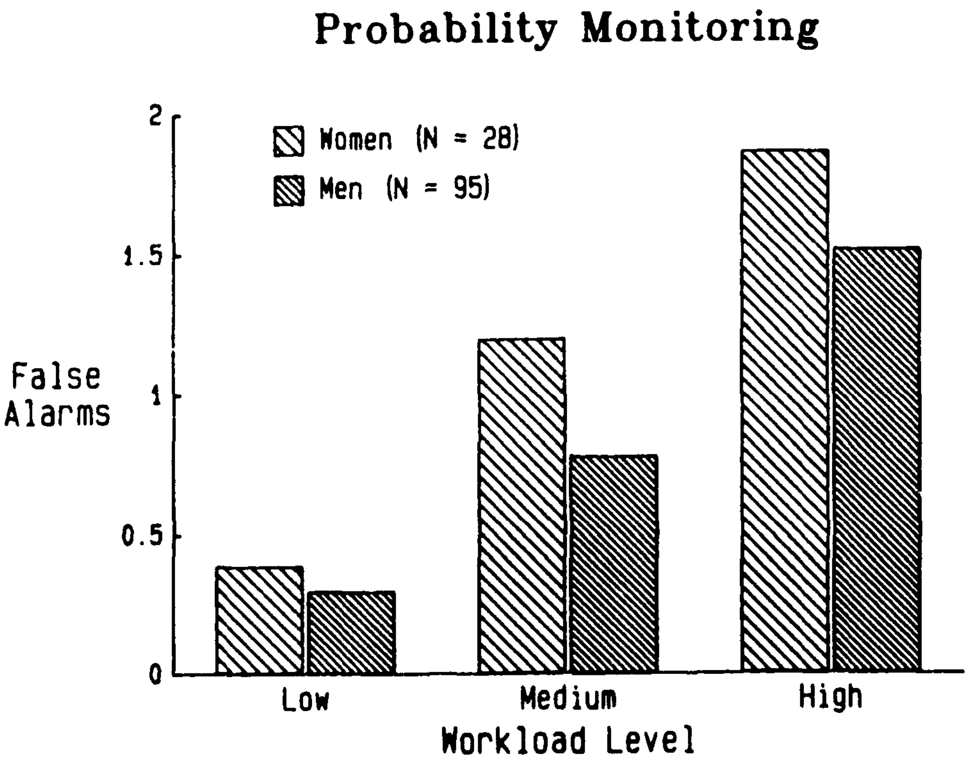


Figure 4-43. False Alarms per Trial for Probability Monitoring - Men vs. Women.

### 4.7.3 Training Data

The means of the major performance measures for Probability Monitoring for training Trials 1 through 5 and baseline Trials 6 and 8 are presented by difficulty level in Table 4-39. Response time, percentage of correct detections, and false alarms are plotted in Figures 4-44 through 4-46.

The analysis of variance model presented in Section 4.1.1 was used with the raw data to determine significance between trials for the various response variables. The results are presented in Table 4-40. Although not significant ($p > 0.22$), there was a steady *increase* of approximately 1 second in response time across the five training trials. This reflects a shift toward more conservative responses in which subjects hesitated longer before claiming that a signal had occurred. This is most evident in the extremely significant decrease ($p < 0.0001$) in the number of false alarms during training (from an average of 3 per trial for Trial 1 to 1 per trial for Trial 5). Unfortunately, this also resulted in a slight decrease ($p \approx 0.02$) in the proportion of correct detections as training progressed, particularly at the high level (from 0.52 for Trial 1 to 0.38 for Trial 5).

Due to the significant Trial by Level interaction for PMPC and PMFA, separate analyses were performed for each level using a reduced model involving only the trial and subject effects. The results of Tukey studentized range tests are summarized in Table 4-41.

A second set of analyses were performed by collapsing the data across subjects prior to the Level by Trial analysis. Data for all five training trials plus the two baseline trials were analyzed. Significant differences among trials were found for PMRT ($p = 0.04$) and PMFA ($p < 0.0001$). Again the results showed that response time worsened but false alarm rate improved over time.

Table 4-39. Means of Performance Measures, Probability Monitoring - Training Trials.

| Var. | Level | Trial | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 8 |
| PMRT | Low<br>Med<br>High | 6.8<br>14.0<br>17.2 | 7.8<br>14.6<br>16.9 | 7.7<br>14.8<br>17.2 | 7.8<br>15.1<br>17.8 | 8.6<br>15.9<br>16.8 | 8.5<br>15.9<br>18.3 | 8.2<br>16.0<br>16.7 |
| PMPC | Low<br>Med<br>High | 0.99<br>0.83<br>0.54 | 0.97<br>0.87<br>0.50 | 0.99<br>0.80<br>0.49 | 0.99<br>0.83<br>0.49 | 0.98<br>0.82<br>0.39 | 0.98<br>0.79<br>0.43 | 0.98<br>0.79<br>0.40 |
| PMFA | Low<br>Med<br>High | 2.01<br>3.67<br>4.09 | 1.06<br>2.15<br>2.67 | 0.58<br>1.46<br>2.29 | 0.47<br>1.37<br>2.14 | 0.43<br>1.06<br>1.93 | 0.41<br>0.86<br>1.54 | 0.24<br>0.89<br>1.66 |

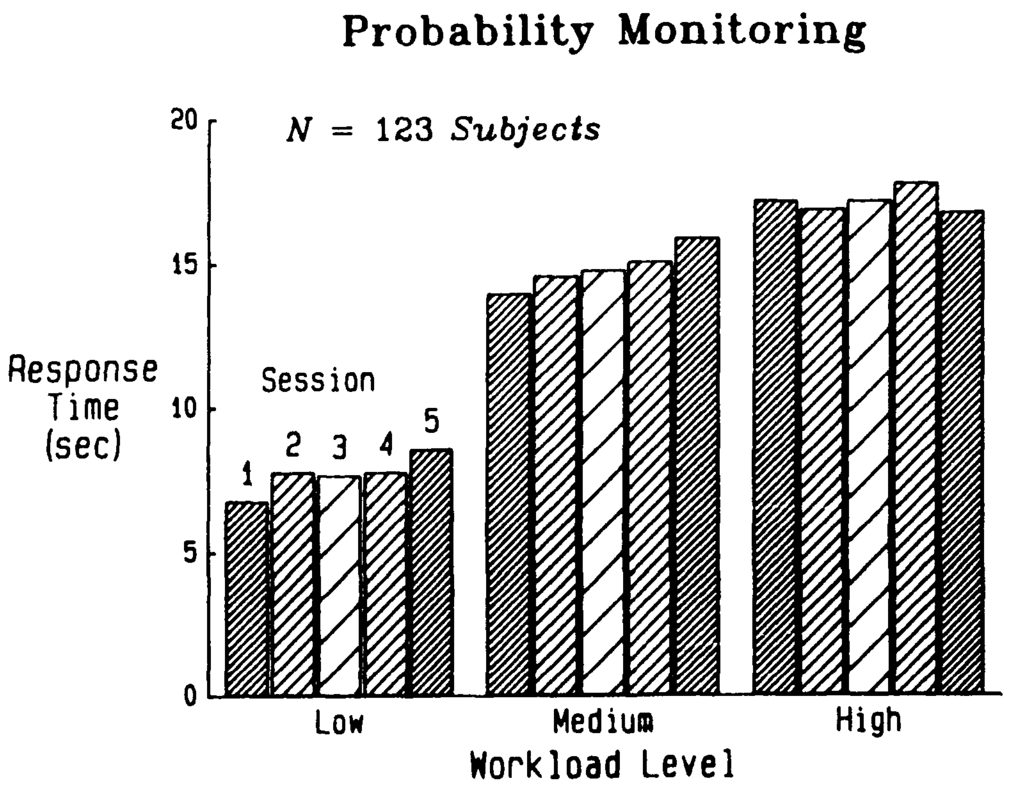


Figure 4-44. Response Time for Probability Monitoring - Trials 1 through 5.

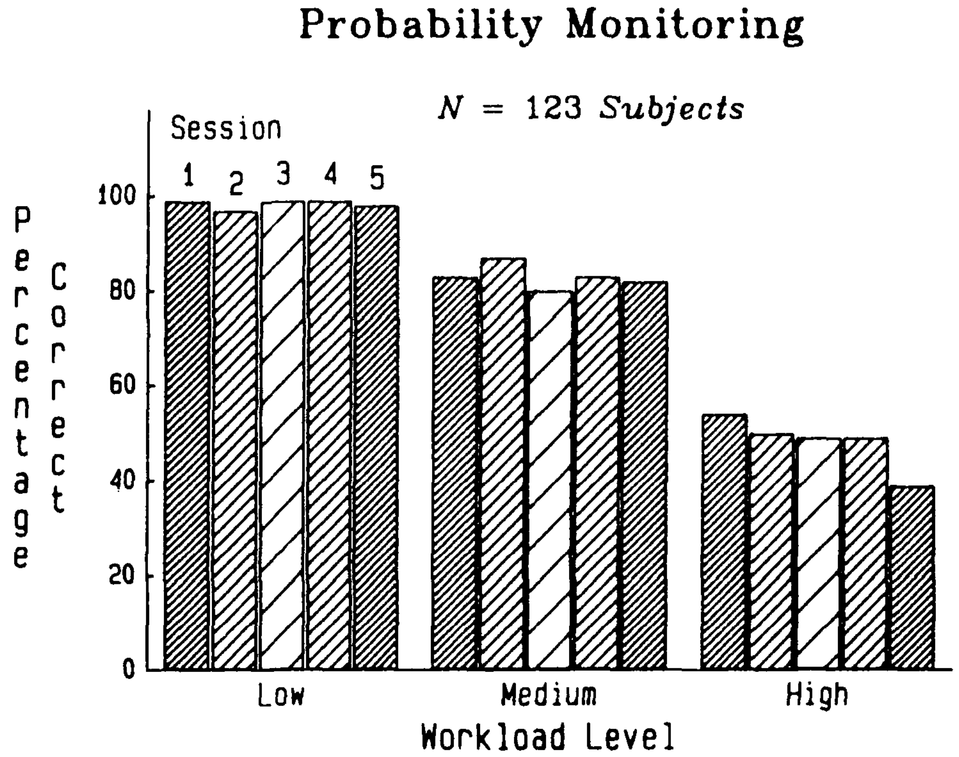


Figure 4-45. Percent Correct for Probability Monitoring - Trials 1 through 5.

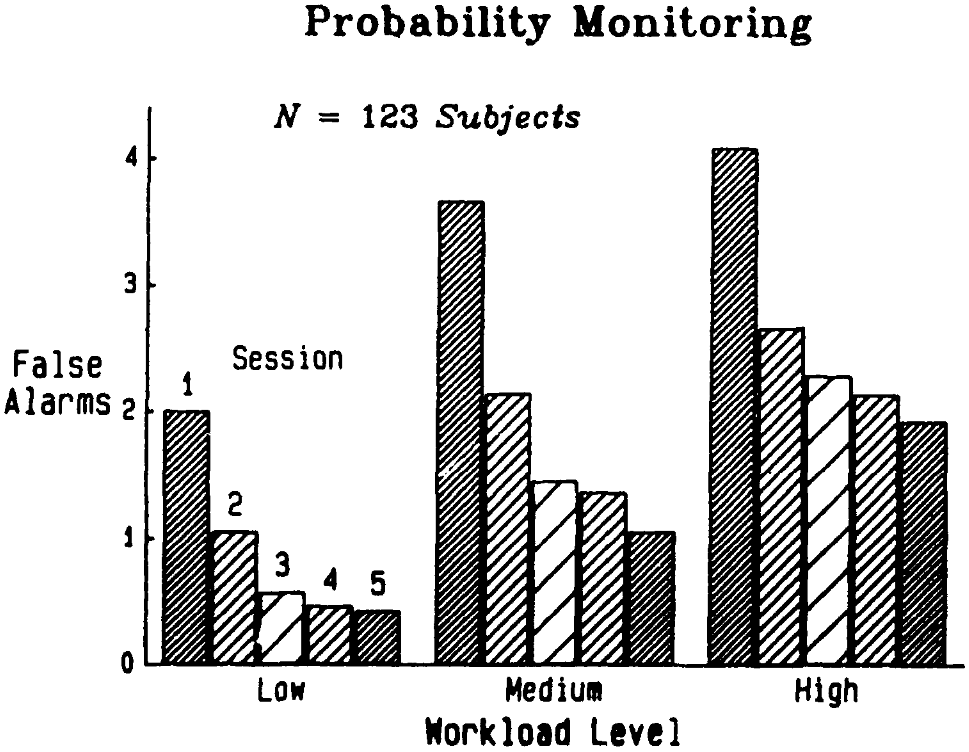


Figure 4-46. False Alarms per Trial for Probability Monitoring - Trials 1 through 5.

Table 4-40. ANOVA Summary for Level and Trial Effects, Probability Monitoring - Training Trials.

| Var. | Model | Level | | Trial | | Level * Trial | |
|---|---|---|---|---|---|---|---|
| | $R^2$ | $F_{(2,244)}$ | $p>F$ | $F_{(4,488)}$ | $p>F$ | $F_{(8,976)}$ | $p>F$ |
| PMRT | 0.73 | 316.29 | * | 1.42 | .2268 | 1.29 | .2448 |
| PMPC | 0.75 | 585.26 | * | 2.98 | .0189 | 2.58 | .0086 |
| PMFA | 0.78 | 113.27 | * | 60.18 | * | 1.82 | .0698 |

* p < 0.0001

Table 4-41. Significant (α = .01) Trial Differences by Level.

| Var. | Level | $F_{(4,488)}$ | Trial |
|---|---|---|---|
| PMRT | L | 7.11 | 5 4 2 3 1 |
| | M | 0.98 | 5 3 4 2 1 |
| | H | 0.47 | 2 5 1 3 4 |
| PMPC | L | 1.61 | 2 5 4 3 1 |
| | M | 1.66 | 3 5 4 1 2 |
| | H | 3.37 | 5 3 4 2 1 |
| PMFA | L | 21.81 | 1 2 3 4 5 |
| | M | 38.17 | 1 2 3 4 5 |
| | H | 23.70 | 1 2 3 4 5 |

### 4.8 Spatial Processing Task

The means and standard deviations for the Spatial Processing performance measures are presented in Table 4-42 for Trials 6 and 8. Overall mean response time (RT) and percentage correct (PC) are presented in Figures 4-47 and 4-48 with the intertrial correlations represented by the *r* values in the figures. Univariate summaries of the overall response time and proportion correct measures for the three difficulty levels are provided in Appendix A-8. Data for these summaries included the Trial 6 and Trial 8 data for all 123 subjects.

Response times were approximately 0.8, 1.3 and 1.5 seconds for the low, medium and high difficulty levels with standard deviations that increased with increasing difficulty from approximately 0.3 to 0.5 seconds. The small difference in response time between the medium and high levels was also reflected in comments by some subjects that the high level (6 bars, 180°) seemed easier than the medium level (4 bars, 90° or 270°). However, proportion correct decreased steadily from 0.95 (low) to 0.92 (medium) to 0.90 (high) with an accompanying increase in standard deviation from 0.05 to 0.09.

Response times were essentially equal for *MATCH* (right button) and *NON-MATCH* (left button) responses. However, the proportion correct was noticeably higher for *MATCH* responses at the medium (0.96 vs. 0.88) and high (0.94 vs. 0.85) difficulty levels. The larger drop in accuracy across levels for the *NON-MATCH* responses was the primary cause for the decrease in the overall proportion correct.

The *r* values for response time were large (0.74 to 0.91), indicating high stability in performance for this task. The consistently low values for percentage correct (0.33 to 0.38) were again probably due to the high-accuracy ceiling effect.

#### 4.8.1 Level and Trial Analyses

Analysis of variance was performed to verify the difficulty level manipulation and examine any trial differences using the model presented in Section 4.1.1. The results of the analyses for response time and proportion correct are summarized in Table 4-43. The model $R^2$ values were 0.96 for RT and 0.84 for PC. The Tukey studentized range test at $\alpha = 0.01$ demonstrated that all three difficulty levels differed significantly for both RT and PC.

The mean response time for Trial 8 was statistically lower (p < 0.0001) than the RT for Trial 6 particularly at the medium and high levels. However, the average difference amounted to 80 msec (~ 6%). The proportion correct values were identical for Trials 6 and 8.

Table 4-42. Means (Standard Deviations) of Performance Measures, Spatial Processing - Baseline Trials.

| Level | Low | | | Medium | | | High | | |
|---|---|---|---|---|---|---|---|---|---|
| Trial | 06 | 08 | Both | 06 | 08 | Both | 06 | 08 | Both |
| MNO | 768<br>(247) | 748<br>(257) | 758<br>(252) | 1344<br>(414) | 1234<br>(374) | 1289<br>(397) | 1593<br>(522) | 1485<br>(493) | 1539<br>(510) |
| SDO | 296<br>(221) | 319<br>(309) | 308<br>(269) | 612<br>(424) | 551<br>(361) | 582<br>(394) | 734<br>(383) | 638<br>(336) | 686<br>(363) |
| PCO | 95<br>(5) | 95<br>(5) | 95<br>(5) | 92<br>(7) | 92<br>(7) | 92<br>(7) | 89<br>(9) | 90<br>(8) | 90<br>(9) |
| STIMO | 26<br>(1) | 26<br>(1) | 26<br>(1) | 24<br>(1) | 24<br>(1) | 24<br>(1) | 23<br>(2) | 23<br>(2) | 23<br>(2) |
| MNP | 762<br>(268) | 733<br>(250) | 747<br>(259) | 1368<br>(425) | 1262<br>(440) | 1315<br>(435) | 1617<br>(553) | 1491<br>(538) | 1554<br>(548) |
| SDP | 267<br>(241) | 279<br>(245) | 273<br>(243) | 535<br>(424) | 466<br>(339) | 500<br>(385) | 609<br>(376) | 509<br>(273) | 559<br>(332) |
| PCP | 96<br>(7) | 95<br>(7) | 95<br>(7) | 96<br>(7) | 96<br>(7) | 96<br>(7) | 93<br>(9) | 95<br>(8) | 94<br>(9) |
| MNN | 775<br>(262) | 771<br>(338) | 773<br>(302) | 1354<br>(529) | 1207<br>(404) | 1281<br>(476) | 1580<br>(582) | 1485<br>(523) | 1533<br>(554) |
| SDN | 273<br>(246) | 316<br>(499) | 295<br>(393) | 639<br>(535) | 540<br>(455) | 589<br>(498) | 753<br>(486) | 668<br>(465) | 710<br>(476) |
| PCN | 94<br>(8) | 94<br>(7) | 94<br>(8) | 87<br>(13) | 88<br>(12) | 88<br>(13) | 85<br>(16) | 85<br>(14) | 85<br>(15) |

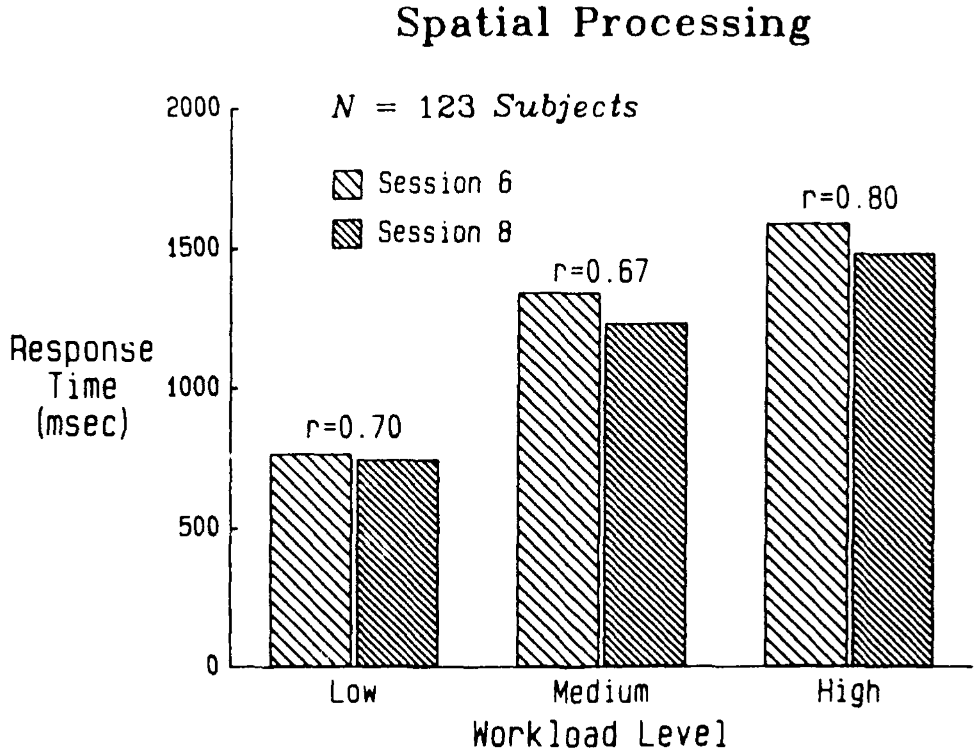


**Figure 4-47. Mean Response Time for Spatial Processing - Trials 6 and 8.**

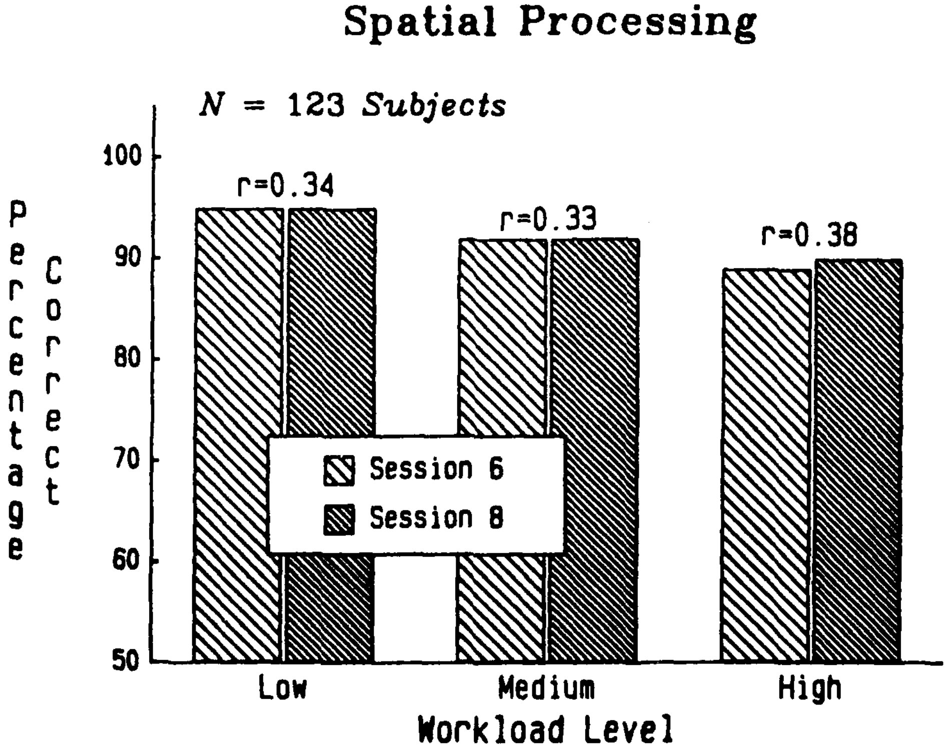


**Figure 4-48. Percentage Correct for Spatial Processing - Trials 6 and 8.**

Table 4-43. ANOVA Summary for Level and Trial Effects, Spatial Processing - Baseline Trials.

**Response Time**

| Source | DF | Sum of Squares (x1000) | Mean Square (x1000) | F | p > F |
|---|---|---|---|---|---|
| Level (L) | 2 | 78226 | 39113 | 396.08 | 0.0001 |
| Trial (T) | 1 | 1162 | 1162 | 19.36 | 0.0001 |
| L by T | 2 | 322 | 161 | 5.18 | 0.0062 |
| Subject (S) | 122 | 77400 | 634 | 20.45 | 0.0001 |
| L by S | 244 | 24095 | 99 | 3.18 | 0.0001 |
| T by S | 122 | 7322 | 60 | 1.93 | 0.0001 |
| Error | 244 | 7570 | 31 | | |

**Proportion Correct**

| Source | DF | Sum of Squares (x0.001) | Mean Square (x0.001) | F | p > F |
|---|---|---|---|---|---|
| Level (L) | 2 | 337 | 168 | 33.33 | 0.0001 |
| Trial (T) | 1 | 3 | 3 | 0.62 | 0.4317 |
| L by T | 2 | 5 | 3 | 0.90 | 0.4094 |
| Subject (S) | 122 | 1357 | 11 | 3.99 | 0.0001 |
| L by S | 244 | 1232 | 5 | 1.81 | 0.0001 |
| T by S | 122 | 545 | 4 | 1.60 | 0.0010 |
| Error | 244 | 681 | 3 | | |

### 4.8.2 Gender and Prototype Analyses

As described in Section 4.1.2 for CR, further analyses of gender and prototype differences were performed. The performance measures are presented separately for men and women in Table 4-44 and Figures 4-49 and 4-50. There were no differences between men and women for response time or proportion correct although there was a significant Trial by Gender interaction (p = 0.0004) for response time.

There were no significant differences among the prototype groups for either response variable. Refer to Section 5.2 for a discussion of the prototype grouping.

## Table 4-44. Means (Standard Deviations) of Performance Measures by Gender, Spatial Processing - Baseline Trials.

| Level | Low | | | Medium | | | High | | |
|---|---|---|---|---|---|---|---|---|---|
| Gender | Fem | Male | Both | Fem | Male | Both | Fem | Male | Both |
| MNO | 782<br>(316) | 751<br>(230) | 758<br>(252) | 1262<br>(406) | 1297<br>(396) | 1289<br>(397) | 1569<br>(568) | 1530<br>(492) | 1539<br>(510) |
| SDO | 327<br>(378) | 302<br>(228) | 308<br>(269) | 550<br>(321) | 591<br>(413) | 582<br>(394) | 681<br>(353) | 688<br>(366) | 686<br>(363) |
| PCO | 96<br>(5) | 94<br>(5) | 95<br>(5) | 93<br>(6) | 92<br>(8) | 92<br>(7) | 88<br>(11) | 90<br>(8) | 90<br>(9) |
| STIMO | 26<br>(1) | 26<br>(1) | 26<br>(1) | 24<br>(1) | 24<br>(1) | 24<br>(1) | 23<br>(2) | 23<br>(1) | 23<br>(2) |
| MNP | 760<br>(296) | 744<br>(248) | 747<br>(259) | 1273<br>(416) | 1328<br>(441) | 1315<br>(435) | 1574<br>(555) | 1548<br>(548) | 1554<br>(548) |
| SDP | 232<br>(191) | 286<br>(255) | 273<br>(243) | 503<br>(349) | 499<br>(396) | 500<br>(385) | 529<br>(293) | 568<br>(342) | 559<br>(332) |
| PCP | 95<br>(7) | 95<br>(7) | 95<br>(7) | 97<br>(6) | 96<br>(8) | 96<br>(7) | 95<br>(9) | 94<br>(9) | 94<br>(9) |
| MNN | 813<br>(438) | 762<br>(248) | 773<br>(302) | 1288<br>(563) | 1278<br>(448) | 1281<br>(476) | 1569<br>(657) | 1522<br>(523) | 1533<br>(554) |
| SDN | 363<br>(693) | 274<br>(243) | 295<br>(393) | 533<br>(332) | 606<br>(536) | 589<br>(498) | 747<br>(507) | 700<br>(468) | 710<br>(476) |
| PCN | 96<br>(5) | 93<br>(8) | 94<br>(8) | 89<br>(13) | 87<br>(13) | 88<br>(13) | 83<br>(20) | 86<br>(13) | 85<br>(15) |

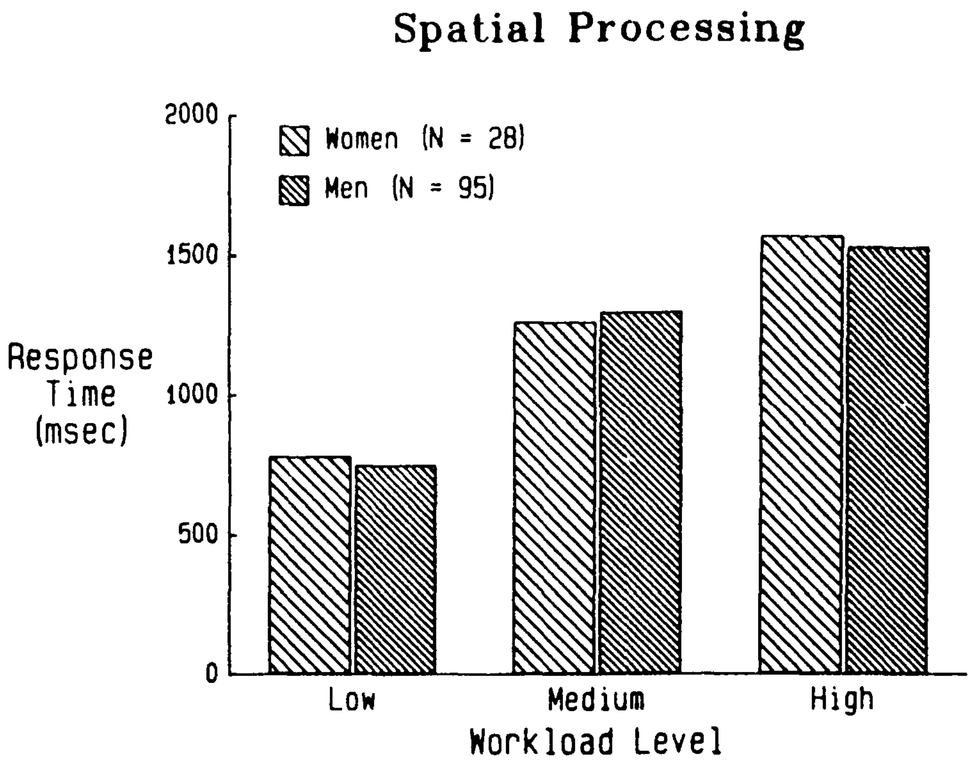


Figure 4-49. Mean Response Time for Spatial Processing - Men vs. Women.

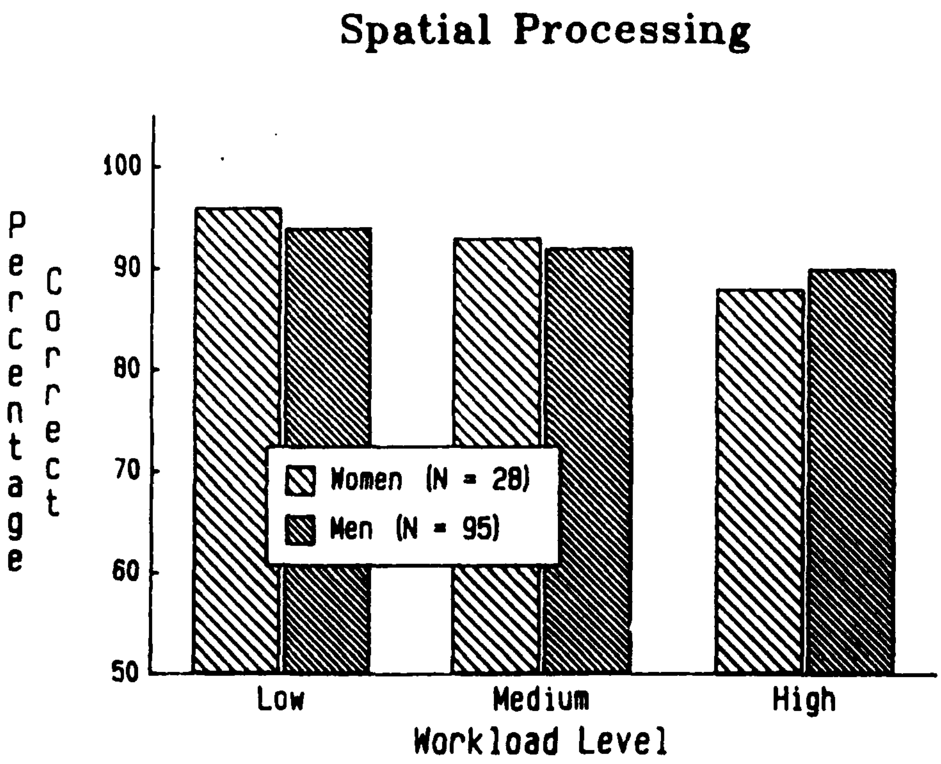


Figure 4-50. Percentage Correct for Spatial Processing - Men vs. Women.

### 4.8.3 Training Data

The means and standard deviations of the major performance measures for Spatial Processing for training Trials 1 through 5 are presented by difficulty level in Table 4-45. Response time and percentage correct are plotted in Figures 4-51 and 4-52. There was a steady improvement in speed and a slight improvement in accuracy during training for the SP task. The largest change in response time was from Trial 1 to Trial 2.

Analysis of variance was used to determine significance between trials for RT and PC using the model presented in Section 4.1.1. A summary of the ANOVA results is presented in Table 4-46. Due to the significant Trial by Level interaction for PC, separate analyses were performed for each level using a reduced model involving only the trial and subject effects. The results of Tukey studentized range tests are summarized in Table 4-47. With respect to RT, there were no significant differences among Trials 2 through 5 at the low and high levels or among Trials 3, 4 and 5 at the medium level with no improvement beyond Trial 4. For PC, there were no differences among all five trials except at the high level where Trial 1 demonstrated slightly lower accuracy than Trials 4 and 5.

Table 4-45. Means (Standard Deviations) of Performance Measures, Spatial Processing - Training Trials.

| Var. | Level | Trial | | | | |
|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 |
| MNO | Low | 981 (274) | 822 (201) | 783 (206) | 766 (193) | 776 (246) |
| | Med | 1540 (370) | 1436 (360) | 1358 (372) | 1305 (376) | 1329 (383) |
| | High | 1773 (481) | 1652 (463) | 1606 (466) | 1552 (451) | 1612 (479) |
| SDO | Low | 437 (397) | 296 (168) | 288 (189) | 287 (161) | 289 (191) |
| | Med | 625 (306) | 625 (275) | 558 (338) | 530 (272) | 566 (277) |
| | High | 748 (305) | 696 (277) | 699 (321) | 673 (299) | 742 (379) |
| PCO | Low | 94 (6) | 95 (5) | 95 (5) | 96 (5) | 95 (6) |
| | Med | 90 (7) | 90 (7) | 90 (7) | 92 (7) | 92 (7) |
| | High | 85 (9) | 87 (9) | 87 (10) | 88 (11) | 89 (9) |
| STIMO | Low | 25 (1) | 26 (1) | 26 (1) | 26 (1) | 26 (1) |
| | Med | 23 (1) | 23 (1) | 24 (1) | 24 (1) | 24 (1) |
| | High | 22 (1) | 23 (1) | 23 (1) | 23 (1) | 23 (1) |
| MNP | Low | 969 (328) | 809 (230) | 781 (231) | 753 (210) | 787 (296) |
| | Med | 1568 (430) | 1462 (404) | 1392 (407) | 1339 (424) | 1368 (412) |
| | High | 1772 (498) | 1640 (513) | 1625 (532) | 1577 (519) | 1618 (512) |
| MNN | Low | 982 (286) | 840 (219) | 790 (220) | 783 (210) | 761 (216) |
| | Med | 1508 (414) | 1417 (436) | 1338 (436) | 1276 (413) | 1300 (430) |
| | High | 1800 (566) | 1702 (554) | 1617 (548) | 1534 (480) | 1606 (569) |

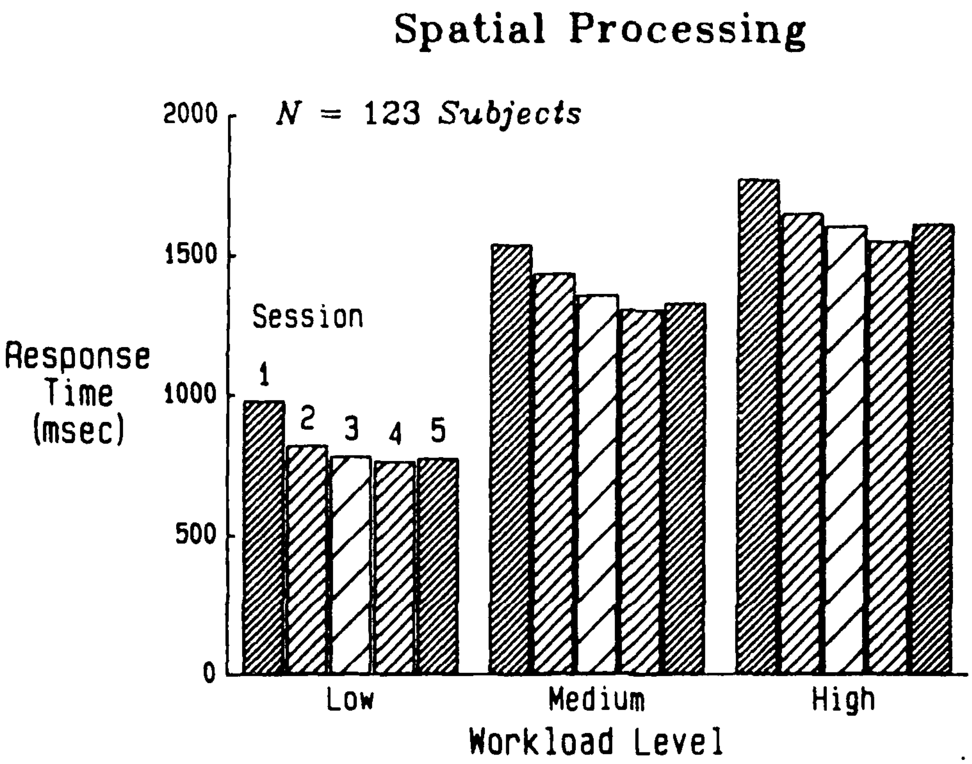


Figure 4-51. Mean Response Time for Spatial Processing - Trials 1 through 5.

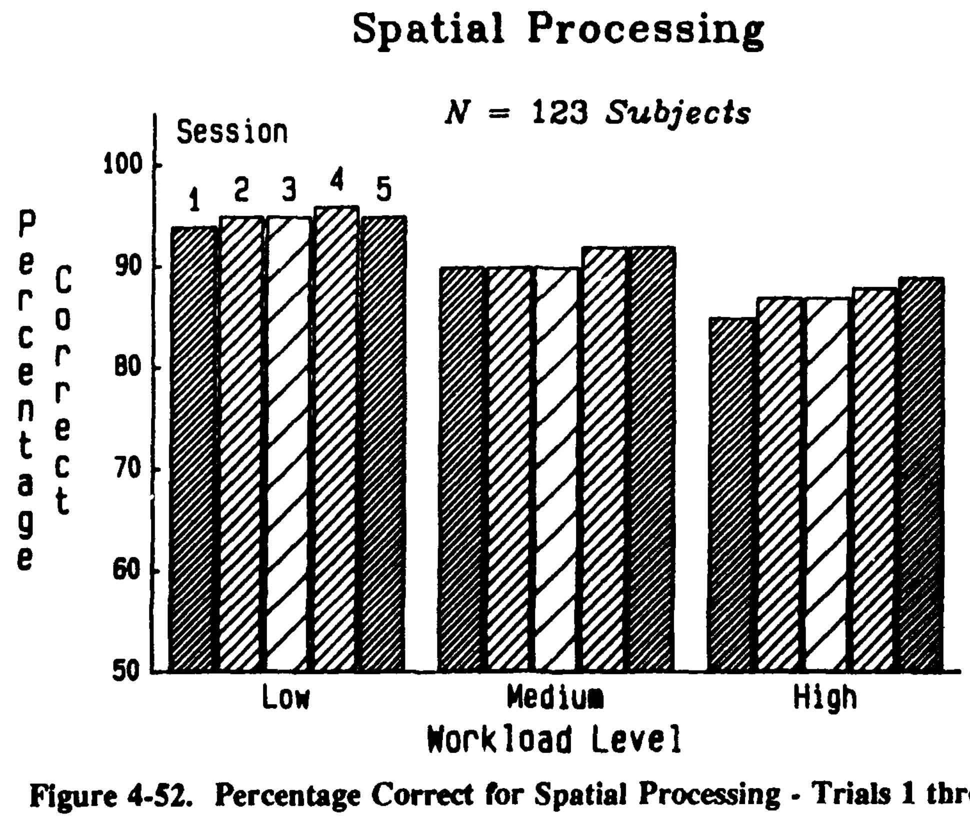


Figure 4-52. Percentage Correct for Spatial Processing - Trials 1 through 5.

Table 4-46. ANOVA Summary for Level and Trial Effects, Spatial Processing - Training Trials.

| Var. | Model | Level | | Trial | | Level * Trial | |
|---|---|---|---|---|---|---|---|
| | $R^2$ | $F_{(2,244)}$ | $p>F$ | $F_{(4,488)}$ | $p>F$ | $F_{(8,976)}$ | $p>F$ |
| SPMNO | 0.92 | 691.23 | * | 33.52 | * | 0.94 | .4819 |
| SPPCO | 0.74 | 91.81 | * | 8.88 | * | 2.12 | .0318 |

* $p < 0.0001$

Table 4-47. Significant ($\alpha = .01$) Trial Differences by Level.

| Var. | Level | $F_{(4,488)}$ | Trial | | | | |
|---|---|---|---|---|---|---|---|
| MNO | L | 51.54 | 1 | 2 | 3 | 5 | 4 |
| | M | 17.92 | 1 | 2 | 3 | 5 | 4 |
| | H | 10.24 | 1 | 2 | 5 | 3 | 4 |
| PCO | L | 2.80 | 1 | 5 | 2 | 3 | 4 |
| | M | 3.19 | 1 | 3 | 2 | 4 | 5 |
| | H | 6.36 | 1 | 2 | 3 | 4 | 5 |

## 4.9 Unstable Tracking Task

The means and standard deviations for the Unstable Tracking performance measures are presented in Table 4-48 for Trials 6 and 8. Mean absolute error (UTMAE) and number of edge violations (UTEV) are presented in Figures 4-53 and 4-54 with the intertrial correlations represented by the $r$ values in the figures. As mentioned in Section 3.4, all summaries and analyses were based on 120 subjects since the edge violation scores for one female subject (#5) and two male subjects (#72, #93) were determined to be outliers. Univariate summaries of the performance measures based on the Trial 6 and Trial 8 data for the 120 subjects are provided in Appendix A-9.

Mean absolute error scores were 10, 33 and 37 for the low, medium and high difficulty levels. In contrast with all other CTS tasks, the standard deviations for UTMAE *decreased* with increasing difficulty from 9 to 6. This may indicate an inherent limiting characteristic of the task itself or of this particular scoring measure. Due to the physical size limits (and number of scan lines) of the video display, there is an upper limit to the maximum error during any one-second period.

Edge violations increased from approximately 6 to 150 to 406 across the three difficulty levels, with extremely large increases in standard deviation from 17 to 132 to 176.

The $r$ values for UTMAE were among the highest of all the tasks (0.81 to 0.91), indicating excellent stability. The values for UTEV were moderately high (0.63 to 0.82), again indicating good stability for the UT task.

### 4.9.1 Level and Trial Analyses

Analysis of variance was performed to verify the difficulty level manipulation and examine any trial differences using the model presented in Section 4.1.1. The results of the analyses for UTMAE and UTEV are summarized in Table 4-49. The model $R^2$ values were extremely high at 0.99 for UTMAE and 0.98 for UTEV. The Tukey studentized range test at $\alpha = 0.01$ demonstrated that all three difficulty levels differed significantly for both UTMAE and UTEV.

Tracking error was significantly lower for Trial 8 than for Trial 6 as measured by UTMAE ($p < 0.0001$) and UTEV ($p = 0.001$). There was a significant ($p = 0.02$) Level by Trial interaction for UTEV only.

Table 4-48. Means (Standard Deviations) of Performance Measures, Unstable Tracking - Baseline Trials.

| Level | Low | | | Medium | | | High | | |
|---|---|---|---|---|---|---|---|---|---|
| Trial | 06 | 08 | Both | 06 | 08 | Both | 06 | 08 | Both |
| UTMAE | 11.0<br>(10.0) | 9.4<br>(8.2) | 10.2<br>(9.2) | 33.1<br>(7.3) | 32.1<br>(7.5) | 32.6<br>(7.4) | 37.0<br>(6.1) | 36.2<br>(6.0) | 36.6<br>(6.0) |
| UTEV | 8.2<br>(20.8) | 4.6<br>(11.1) | 6.4<br>(16.7) | 159.9<br>(138.1) | 140.0<br>(125.8) | 150.0<br>(132.2) | 422.0<br>(184.0) | 390.0<br>(167.3) | 406.0<br>(176.2) |

Table 4-49. ANOVA Summary for Level and Trial Effects, Unstable Tracking - Baseline Trials.

| Mean Absolute Error | | | | | |
|---|---|---|---|---|---|
| Source | DF | Sum of Squares | Mean Square | F | p > F |
| Level (L) | 2 | 97065 | 48533 | 728.48 | 0.0001 |
| Trial (T) | 1 | 234 | 234 | 17.05 | 0.0001 |
| L by T | 2 | 19 | 9 | 1.62 | 0.2004 |
| Subject (S) | 119 | 22723 | 191 | 32.59 | 0.0001 |
| L by S | 238 | 15856 | 67 | 11.37 | 0.0001 |
| T by S | 119 | 1636 | 14 | 2.35 | 0.0001 |
| Error | 238 | 1394 | 6 | | |
| Edge Violations | | | | | |
| Source | DF | Sum of Squares (x1000) | Mean Square (x1000) | F | p > F |
| Level (L) | 2 | 19667 | 9833 | 605.85 | 0.0001 |
| Trial (T) | 1 | 62 | 62 | 10.79 | 0.0013 |
| L by T | 2 | 24 | 12 | 4.08 | 0.0181 |
| Subject (S) | 119 | 6324 | 53 | 17.77 | 0.0001 |
| L by S | 238 | 3863 | 16 | 5.43 | 0.0001 |
| T by S | 119 | 679 | 6 | 1.91 | 0.0001 |
| Error | 238 | 712 | 3 | | |

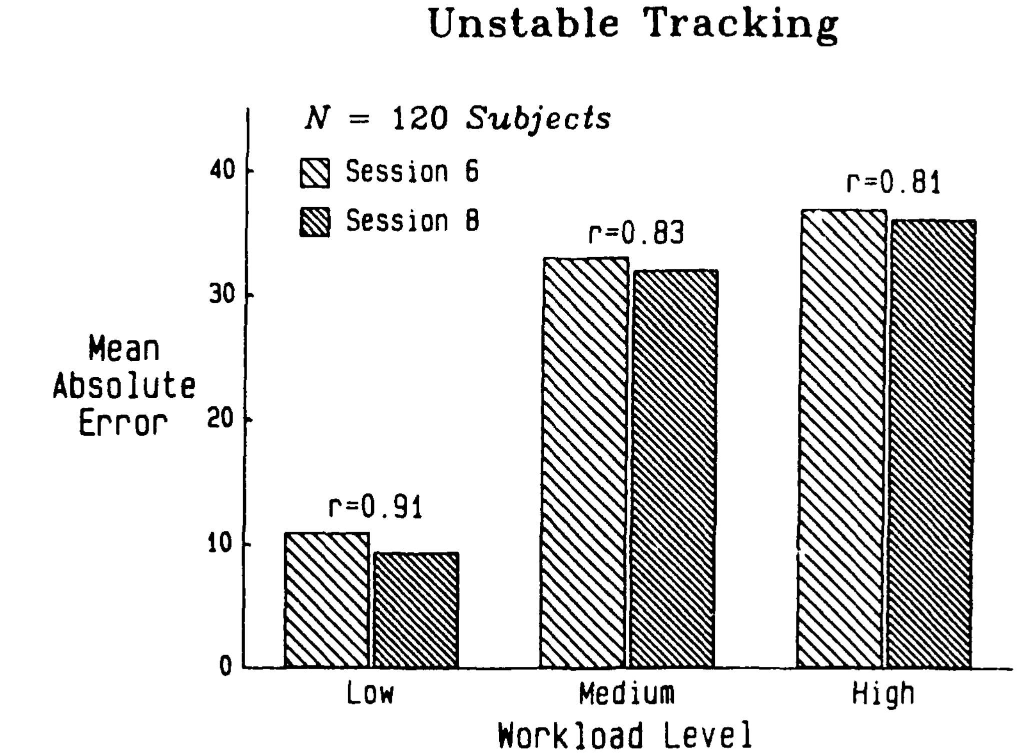


**Figure 4-53. Mean Absolute Error for Unstable Tracking - Trials 6 and 8.**

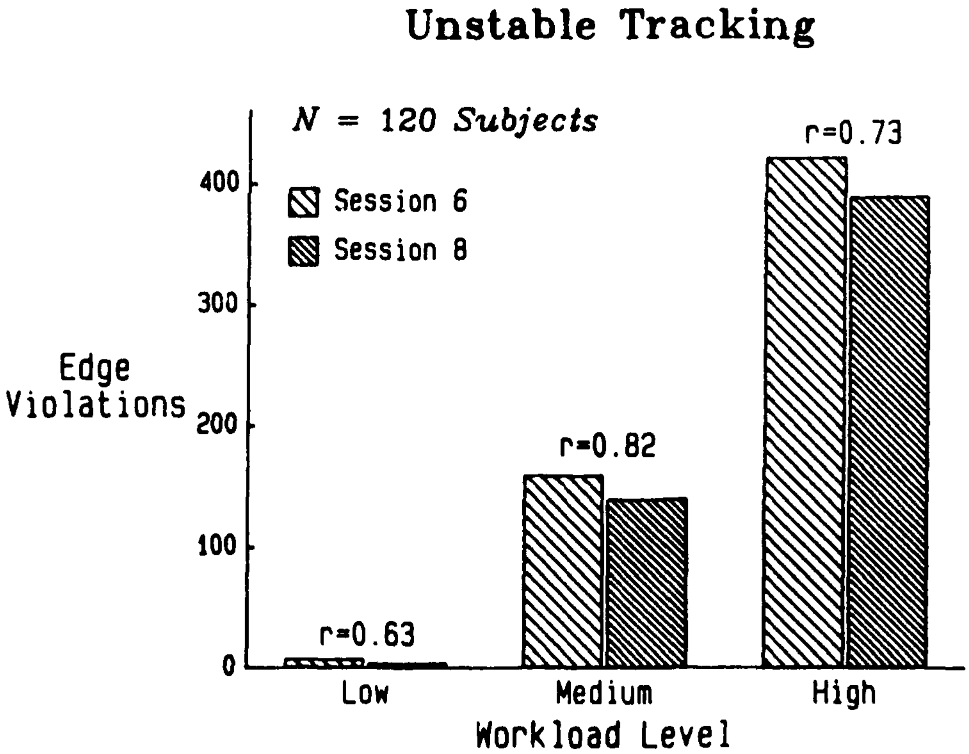


**Figure 4-54. Edge Violations for Unstable Tracking - Trials 6 and 8.**

### 4.9.2 Gender and Prototype Analyses

As described in Section 4.1.2 for CR, further analyses of gender and prototype differences were performed. The performance measures are presented separately for men and women in Table 4-50 and Figures 4-55 and 4-56. Men provided better tracking performance than women with a significant difference in scores for UTMAE ($p < 0.05$) and a marginally significant difference for UTEV ($p < 0.10$). There was a marginally significant Level by Gender interaction for UTEV due to the smaller magnitude of the difference at the low level. On a percentage basis, however, the number of edge violations for males was 59%, 67% and 93% of the number for females at the low, medium and high levels respectively.

There was no significant difference in tracking performance among the prototype groups, with essentially identical scores on UTMAE for all four groups and slight differences in scores for UTEV. Refer to Section 5.2 for a discussion of the prototype grouping.

**Table 4-50. Means (Standard Deviations) of Performance Measures by Gender, Unstable Tracking - Baseline Trials.**

| Level | Low | | | Medium | | | High | | |
|---|---|---|---|---|---|---|---|---|---|
| Gender | Fem | Male | Both | Fem | Male | Both | Fem | Male | Both |
| UTMAE | 13.0<br>(9.9) | 9.4<br>(8.8) | 10.2<br>(9.2) | 34.9<br>(6.6) | 31.9<br>(7.5) | 32.6<br>(7.4) | 37.1<br>(6.1) | 36.4<br>(6.0) | 36.6<br>(6.0) |
| UTEV | 9.4<br>(18.4) | 5.5<br>(16.2) | 6.4<br>(16.7) | 200.9<br>(138.6) | 135.2<br>(126.9) | 150.0<br>(132.2) | 428.7<br>(190.5) | 399.4<br>(171.8) | 406.0<br>(176.2) |

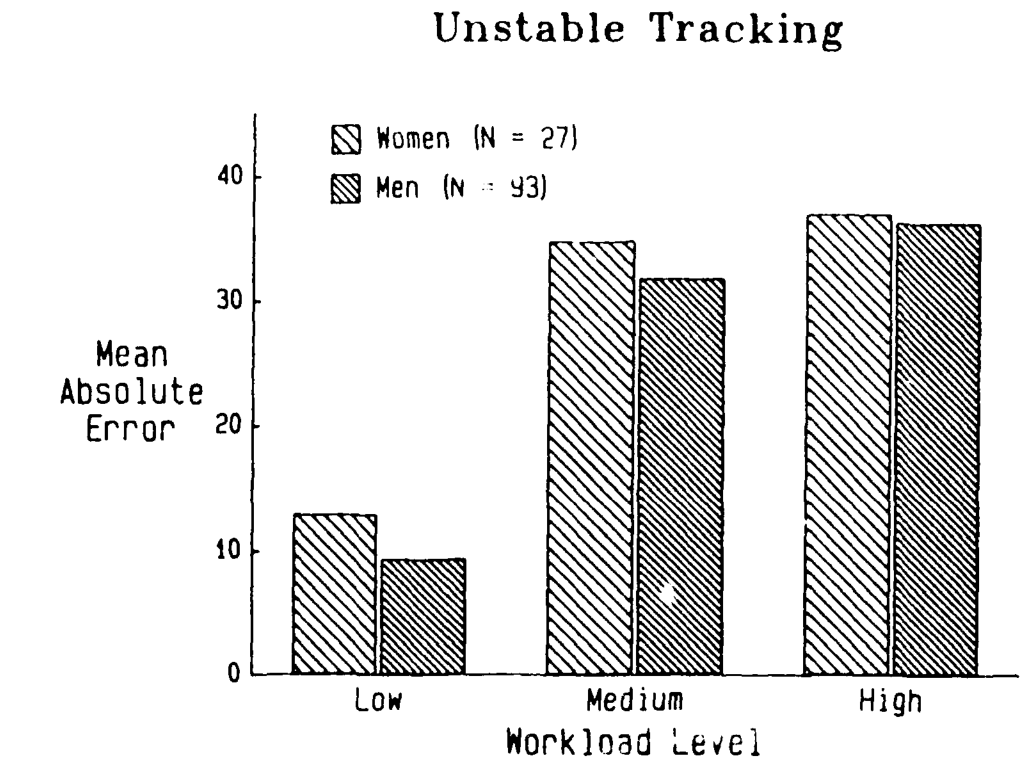


Figure 4-55. Mean Absolute Error for Unstable Tracking - Men vs. Women.

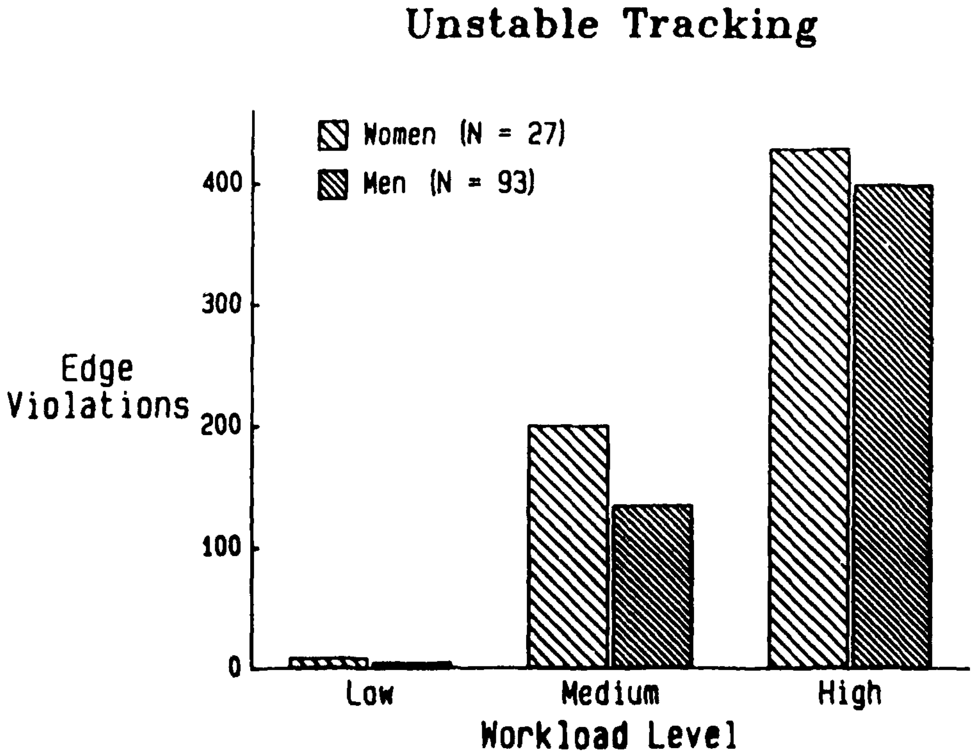


Figure 4-56. Edge Violations for Unstable Tracking - Men vs. Women.

### 4.9.3 Training Data

The means and standard deviations of the major performance measures for Unstable Tracking for training Trials 1 through 5 are presented by difficulty level in Table 4-51. Mean absolute error and edge violations are plotted in Figures 4-57 and 4-58. The steady improvement in tracking performance was most evident in UTEV which continued to increase throughout all five trials. The largest improvement occurred between the first and second trial. UTMAE also demonstrated large improvement from Trial 1 to Trial 2 but the improvement leveled off at Trial 3.

Analysis of variance using the model presented in Section 4.1.1 verified highly significant differences and interactions for both UTMAE and UTEV (all at $p < 0.0001$). A summary of the ANOVA results is presented in Table 4-52. Separate analyses were performed for each level using a reduced model involving only the trial and subject effects. The results of Tukey studentized range tests at $\alpha = 0.01$ are summarized in Table 4-53. In general, the results for UTMAE and UTEV were quite similar. With respect to UTMAE, there were no significant differences among Trials 2 through 5 at the low and high levels or between Trials 4 and 5 at the medium level. For UTEV, there were no differences among Trials 2 through 5 at the low and high levels or among Trials 3, 4 and 5 at the medium level.

**Table 4-51. Means (Standard Deviations) of Performance Measures, Unstable Tracking - Training Trials.**

| Var. | Level | Trial | | | | |
|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 |
| | Low | 16.9 (10.9) | 11.5 (9.1) | 10.7 (8.9) | 10.7 (9.8) | 10.8 (9.2) |
| UTMAE | Med | 40.7 ( 4.2) | 38.0 (4.7) | 36.5 (5.8) | 35.3 (7.3) | 34.8 (6.8) |
| | High | 41.2 ( 7.0) | 39.3 (5.3) | 38.6 (5.7) | 39.0 (7.2) | 38.0 (6.6) |
| | Low | 29.7 ( 50.9) | 8.4 ( 17.8) | 6.3 ( 17.0) | 9.6 ( 29.7) | 7.2 ( 18.1) |
| UTEV | Med | 339.7 (228.1) | 239.3 (126.0) | 218.8 (178.3) | 195.8 (156.6) | 185.9 (154.3) |
| | High | 592.8 (375.2) | 494.5 (213.0) | 477.5 (227.0) | 507.3 (318.1) | 467.1 (260.3) |

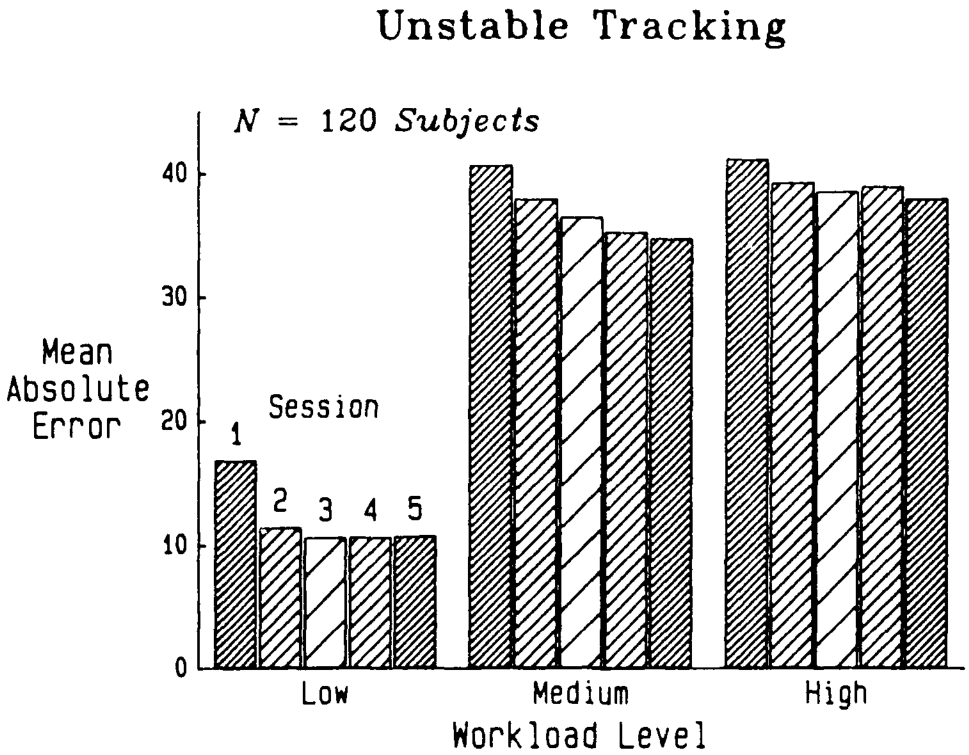


Figure 4-57. Mean Absolute Error for Unstable Tracking - Trials 1 through 5.

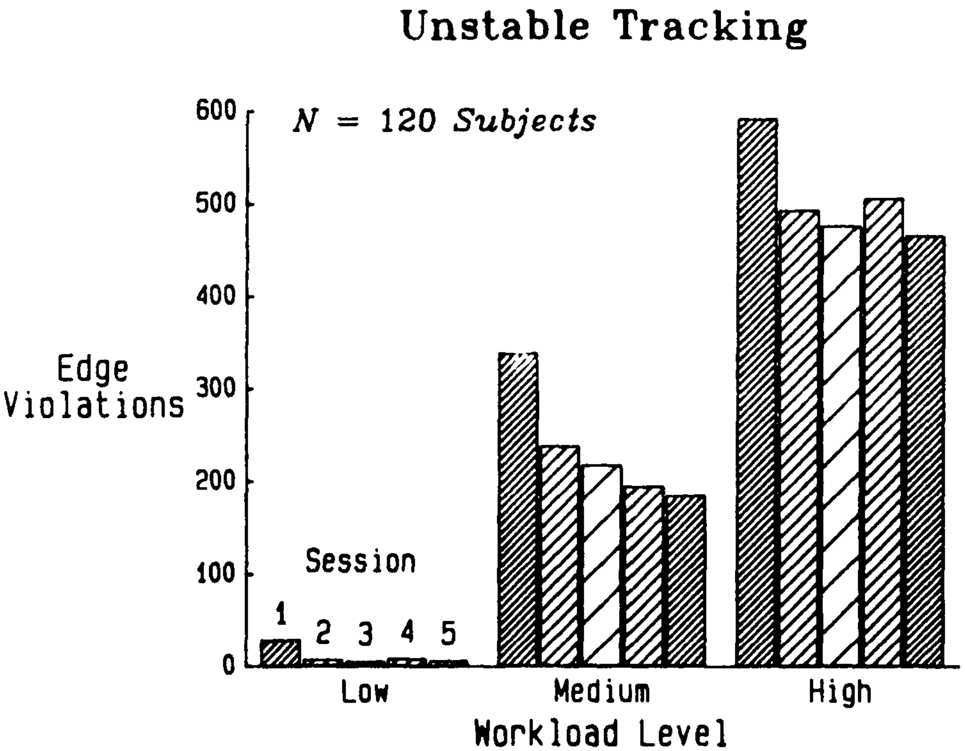


Figure 4-58. Edge Violations for Unstable Tracking - Trials 1 through 5.

Table 4-52. ANOVA Summary for Level and Trial Effects, Unstable Tracking - Training Trials.

| Var. | Model | Level | | Trial | | Level * Trial | |
|---|---|---|---|---|---|---|---|
| | $R^2$ | $F_{(2,238)}$ | $p>F$ | $F_{(4,476)}$ | $p>F$ | $F_{(8,952)}$ | $p>F$ |
| UTMAE | 0.97 | 1445.62 | * | 57.10 | * | 8.10 | * |
| UTEV | 0.91 | 602.93 | * | 20.44 | * | 7.42 | * |

* p < 0.0001

Table 4-53. Significant ($\alpha$ = .01) Trial Differences by Level.

| Var. | Level | $F_{(4,488)}$ | Trial |
|---|---|---|---|
| | L | 40.27 | 1 2 5 4 3 |
| UTMAE | M | 49.15 | 1 2 3 4 5 |
| | H | 9.74 | 1 2 4 3 5 |
| | L | 19.37 | 1 4 2 5 3 |
| UTEV | M | 34.51 | 1 2 3 4 5 |
| | H | 7.32 | 1 4 2 3 5 |

## 4.10 Comparison with Previous Data

Table 4-54 presents a comparison of the data from Trial 8 of the current study with the data reported by Shingledecker (1984). In the current study, the response times were substantially longer for CR and GR (the tasks with the longest times in the present study), somewhat longer for MP and SP, and mixed for LP and MS. Times in the present study were longer at the low and medium levels of LP, but were a full second shorter than the Shingledecker data at the high level. Times for MS were shorter at the low level, equal at the medium level and longer at the high level compared with the 1984 data. The mean CTS Interval Production Variability Score of 29 was within the stated normal range of 10 to 40. Response times for PM were slightly shorter in the current study. Performance on Unstable Tracking as indicated by the Mean Absolute Error was worse in the current study, particularly at the medium level. The number of Edge Violations in the present study was larger at the high level, substantially smaller at the medium level, and about the same at the low level.

**Table 4-54. Comparison of Trial 8 Performance Data with Shingledecker (1984).**

| Level | Low | | | Medium | | | High | | |
|---|---|---|---|---|---|---|---|---|---|
| Task | Current | | 1984 | Current | | 1984 | Current | | 1984 |
| Var. | Mean | Std | Mean | Mean | Std | Mean | Mean | Std | Mean |
| CRMN | 911 | 287 | 550 | 2085 | 870 | 775 | 2971 | 1849 | 1215 |
| GRMN | 3215 | 1117 | 1000 | 5502 | 1497 | 4100 | 7291 | 1844 | 5750 |
| IPVS1 | 29 | 16 | 10-40 | - | - | - | - | - | - |
| LPMN | 517 | 101 | 420 | 772 | 257 | 700 | 1557 | 459 | 2500 |
| MPMN | 540 | 190 | 440 | 1470 | 592 | 1200 | 2528 | 997 | 2030 |
| MSMN | 442 | 76 | 475 | 596 | 137 | 590 | 709 | 155 | 675 |
| PMRT | 8 | - | 12 | 16 | - | 17 | 17 | - | 20 |
| SPMN | 748 | 257 | 550 | 1234 | 374 | 850 | 1485 | 493 | 1250 |
| UTMAE | 9 | 8 | 5 | 32 | 7 | 11 | 36 | 6 | 34 |
| UTEV | 5 | 11 | 5 | 140 | 126 | 225 | 390 | 167 | 280 |

There are at least three probable causes for the performance differences. First, the populations in the two studies are noticeably different. There were fewer subjects in the 1984 study, but these fewer subjects may have represented an overall higher level of motivation and greater experience with the CTS battery and its development. Second, the subjects in the previous study were trained and tested on each task separately rather than with the entire CTS test battery. Finally, subjects in the present study were restricted to five training trials for all tasks and the data that is presented represents an average of the sixth and eighth trials compared with a varying number of training trials for the 1984 study.

These three possibilities may be partially addressed by referring to data from a replication study performed with twenty AFROTC cadets (Schlegel, 1986). The subjects in this study were highly motivated, but had no prior knowledge of the CTS. Five subjects (Group A) performed five trials of all tasks in the battery. The remaining fifteen subjects were split into three groups (Groups B, C and D). Each group performed fifteen trials of one of three different three-task subsets.

For CR, the response times for Group A were substantially longer than those for Group D after five trials. The response times for Group D at Trial 8 were 200 to 500 msec shorter than comparable trials in the current study. However, even after fifteen trials, the times did not approach the faster times reported by Shingledecker.

For GR, the response times for Groups A and B were similar and remarkably close to the data from the current study at Trials 6 and 8. Again, the response times at Trial 15 were substantially longer than in the 1984 study.

The IP Variability Score varied greatly for Group A but improved over time for Group D with values moderately lower than the mean of 29 reported here. In the current study, the score varied greatly rather than improving during the training and baseline trials. As mentioned in Schlegel (1986), it is believed that subjects do not give sufficient attention to the Interval Production task.

For LP, response times for Groups A and D were almost identical and in close agreement with those in the present study at Trials 6 and 8. As in the current study, the times (even at Trial 15) were longer than the 1984 data for the low and medium levels but were much shorter (by 50%) for the high workload level.

For MP, response times for Groups A and C were similar. At Trials 6 and 8, the times were almost identical to those in the current study. At Trial 15, response times for Group C were very close to those reported by Shingledecker. This is one instance of complete agreement across the three studies, with the differences between the current study and the 1984 study explained by differences in the number of training trials. As pointed out by Shingledecker (1984, p. 31), the low difficulty level requires seven training trials, but the medium and high levels require ten training trials. In addition, "performance stability is enhanced if practice is extended to 14 and 30 trials, respectively" on the medium and high levels.

With MS, there is again good agreement among the three studies. The times for Groups A and B were similar, but slightly slower than those at Trials 6 and 8 in the current study. However, times at Trial 15 were quite close to those in the 1984 report.

Response times for PM were still faster in the current study than in the Schlegel (1986) study. However, summary data in the previous studies was highly variable due to the low number of correct detections and the lack of averaging as performed in this study's analysis.

For SP, the response times for Groups A and C were quite similar. Data at Trials 6 and 8 were comparable to data collected in the present study. Response times at Trial 15 were still longer than the Shingledecker data.

With respect to UT Mean Absolute Error, Group C performed better than Group A, but only at the low and medium levels (again pointing out the previously hypothesized limit of this performance variable with increasing λ.) Data in the current study is in approximate agreement with the mean of the Group A and Group C data. Data at Trial 15 is in agreement with the 1984 data only for the high level and is substantially different for the medium level. The Edge Violation scores follow a similar pattern but are still noticeably higher in the current study. At Trial 15, the scores are much lower in the Schlegel (1986) study at the low and medium levels, but are much higher than those in the Shingledecker study at the high level. One explanation for the poorer performance in the current study is the possible individual differences and gradual wear of the potentiometer controllers.

To summarize the above analysis, in only one task (CR) was there disagreement between the data from the present study and the data from the Schlegel (1986) study when compared at Trials 6 and 8. For this task, there was also disagreement between the Schlegel study and the Shingledecker study. In one task (MP), the difference between the data of the present study and that of Shingledecker is attributable to the vastly different number of training trials (5 vs. 15 to 30). In four instances (GR, LP, SP and UT), there was agreement between the current study and the Schlegel (1986) study and common disagreement with the 1984 data, implying that the differences resulted from performing the entire battery vs. isolated tasks, and/or from using naive vs. knowledgeable subjects. In the two remaining instances (MS and PM, tasks requiring less training), there was general agreement among all three studies.

These results also have implications for the central processing Deadlines in the CTS software. The adequacy of these response time constraints is addressed in the Deadline Stressor section of Part II of this report.

### 4.11 Intertask Relationships - Cluster Analysis

A major goal of the current study was to investigate the interrelationships among the CTS tasks and task levels in an attempt to gain a better understanding of the CTS task structure. This would help to determine if all levels of a particular CTS task were drawing from the same resource pool and how much overlap in resource demands existed between tasks.

Several techniques may be employed to determine the underlying structure of a set of measures. These include factor analysis, principle components and multidimensional scaling techniques. The approach selected for this study was clustering analysis. The SAS VARCLUS procedure (*SAS User's Guide: Statistics*, 1985) was used to separate the dependent measures for the various CTS tasks and levels into disjoint clusters. VARCLUS performs the clustering so as to "maximize the sum across clusters of the variance of the original variables that is explained by the cluster components."

A major advantage of cluster analysis is that it can reduce a large set of variables to a set that is more manageable and often easier to interpret. It was used with the CTS data to identify which tasks and task levels provided similar information with respect to resource utilization.

Data from the first baseline trial following training (Trial 6) was analyzed using the Statistical Analysis System VARCLUS procedure to cluster the nine tasks and twenty-five individual task levels. Four separate analyses were performed, three involving the performance data and one with the SWAT data. The SWAT data analysis is reported in Section 5.7.1.

The first clustering analysis included the response time measures for the discrete stimulus tasks, the mean tapping rate (IPMN) and the CTS variability score (IPVS1) for Interval Production, and the Mean Absolute Error (UTMAE) and Edge Violations (UTEV) for Unstable Tracking. Probability Monitoring was excluded from this analysis due to the large number of missing data values for subjects who did not correctly detect any bias signals at some levels. This provided eight variables for each of three levels of seven of the tasks plus two variables for Interval Production for a total of 26 performance measures. Each of the task difficulty levels was included separately in order to determine whether the different levels tap the same resource as indicated by the clustering. A summary of the clusters generated from this analysis is given in Table 4-55.

**Table 4-55. Cluster Analysis for Response Time Measures.**

| Cluster | | | | | | |
|---|---|---|---|---|---|---|
| **1** | **2** | **3** | **4** | **5** | **6** | **7** |
| CR1 | CR2 | GR1 | IPMN | MP1 | SP1 | All |
| | CR3 | GR2 | IPVS | MP2 | SP2 | UT |
| | | GR3 | | MP3 | SP3 | Var |
| LP1 | | | | | | |
| LP2 | | | | | | |
| LP3 | | | | | | |
| MS1 | | | | | | |
| MS2 | | | | | | |
| MS3 | | | | | | |

1 - low level 2 - medium level 3 - high level

Seven clusters of response time variables were identified for the nine tasks. In general, the Memory Search and Linguistic Processing tasks were grouped in one cluster with each of the other clusters representing a single task. This indicated minimal resource overlap for all tasks except these two. The overlap of these two tasks is probably due to the relative ease of the tasks and the similarities of simple symbol manipulation whether linguistic or simple memory update.

With one exception, measures from different levels of the same task were placed in the same cluster indicating that the various workload levels or difficulty manipulations of any given task drew from the same resource pool. The exception was the Continuous Recall task at the low level which was placed in the cluster with LP and MS. This emphasized the much lower difficulty of CR at the low level and its closer association with LP and MS as a symbol manipulation task.

A second clustering analysis examined the accuracy measures of proportion correct in place of the response time measures for the discrete stimulus tasks. Performance measures for the non-central processing tasks were not included in the analysis. The clustering exhibited more task overlap (Table 4-56) than in the analysis of response times. This was probably due to the relatively high level of accuracy for several tasks.

**Table 4-56. Cluster Analysis for Proportion Correct Measures.**

| | | Cluster | | |
|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 |
| CR1 | CR2 | GR1 | MS1 | SP1 |
| | CR3 | GR2 | MS2 | SP2 |
| MP1 | | GR3 | MS3 | SP3 |
| MP2 | | LP2 | LP1 | |
| MP3 | | LP3 | | |

1 - low level   2 - medium level   3 - high level

A final clustering analysis with the performance data combined the response time and accuracy measures for a total of 44 variables. Ten clusters were defined as shown in Table 4-57. Combining all performance measures in a single analysis produced greater overlap of tasks among clusters. However, most measures formed logical clusters along the lines of the first two analyses.

The results of the cluster analysis help to validate the design goals of the Criterion Task Set. The primary area of concern is the overlap among the Linguistic Processing and Memory Search tasks, which both involve somewhat simple symbol manipulation.

**Table 4-57. Cluster Analysis for Performance Measures.**

| Cluster | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| CRMN1 | CRMN2 | CRPC1 | | | CRPC3 | | MPMN1 | SPMN1 | All |
| | CRMN3 | | | | | | MPMN2 | SPMN2 | LT |
| | CRPC2 | | | | IPVS | IPMN | MPMN3 | SPMN3 | Vars |
| LPMN2 | LPPC3 | LPPC2 | LPMN3 | LPPC1 | | | | LPMN1 | |
| | | MPPC1 | | | | | | | |
| | | MPPC2 | | | | | | | |
| | | MPPC3 | | | SPPC2 | | | | |
| | | SPPC1 | | | SPPC3 | | | | |
| MSMN1 | | | | MSPC1 | | | | | |
| MSMN2 | | | | MSPC2 | | | | | |
| MSMN3 | | | | | | MSPC3 | | | |
| | | | GRMN1 | | | GRPC1 | | | |
| | GRPC2 | | GRMN2 | | | | | | |
| | | | GRMN3 | | | | | | |
| | | | GRPC3 | | | | | | |

1 - low level 2 - medium level 3 - high level

## 4.12 Summary of Performance Results

A final summary of the means and standard deviations of the performance variables for all 123 subjects and both baseline trials is provided in Table 4-58.

**Table 4-58. Means (Standard Deviations) of Performance Variables by Task and Level for Trials 6 and 8 Combined (N = 123 Subjects).**

| Task | Level | | | | | |
|---|---|---|---|---|---|---|
| | Low | | Med | | High | |
| | $\overline{x}$ | $(s)$ | $\overline{x}$ | $(s)$ | $\overline{x}$ | $(s)$ |
| CRMN | 960 | (318) | 2105 | (829) | 3077 | (1884) |
| CRPC | 96 | (5) | 87 | (13) | 73 | (13) |
| GRMN | 3252 | (1136) | 5628 | (1514) | 7472 | (1816) |
| GRPC | 93 | (9) | 91 | (11) | 85 | (15) |
| IPMN | 507 | (129) | - | - | - | - |
| IPSD | 52 | (41) | - | - | - | - |
| IPVS1 | 29 | (14) | - | - | - | - |
| IPVS2* | 767 | (312) | - | - | - | - |
| LPMN | 523 | (109) | 792 | (249) | 1578 | (450) |
| LPPC | 97 | (4) | 96 | (3) | 90 | (7) |
| MPMN | 552 | (185) | 1496 | (579) | 2579 | (993) |
| MPPC | 97 | (3) | 97 | (3) | 97 | (5) |
| MSMN | 445 | (71) | 598 | (129) | 726 | (164) |
| MSPC | 97 | (3) | 96 | (5) | 89 | (7) |
| PMRT | 8.4 | ( - ) | 15.9 | ( - ) | 17.5 | ( - ) |
| PMPC | 98 | ( - ) | 79 | ( - ) | 42 | ( - ) |
| PMFA | 0.32 | ( - ) | 0.88 | ( - ) | 1.60 | ( - ) |
| SPMN | 758 | (252) | 1289 | (397) | 1539 | (510) |
| SPPC | 95 | (5) | 92 | (7) | 90 | (9) |
| UTMAE | 10 | (9) | 33 | (7) | 37 | (6) |
| UTEV | 6 | (17) | 150 | (132) | 406 | (176) |

* times a scale factor of 0.0001

## 5.0 SWAT RESULTS

### 5.1 Subjects and SWAT Sorts

Among the 123 subjects who were included in the analyses, there was a varying ability to understand the SWAT and to provide a good initial SWAT sort. Three categories of sorting were identified:

(1) **Good Sorts** - sorts that reflected the subjective ratings of the Time, Effort, and Stress dimensions with few axiom violations and no sort factor reversals, resulting in an acceptable plot of rescaled values vs. raw data,

(2) **Iterative Sorts** - sorts produced by arranging the SWAT combinations in an ascending sequence, thus showing an ability to accurately read the SWAT descriptors and apply logical, mechanistic ordering with little evidence of a subjective evaluation, and

(3) **Poor Sorts** - sorts that indicated a serious lack of understanding of the SWAT descriptors and/or the sorting process.

The SWAT computer analysis was used to identify Category 1 and Category 2 individuals. Category 3 individuals were given a further explanation of SWAT and the sorting process and allowed to complete an additional sort(s). Although this sometimes allowed the subject to move into another category, some individuals remained in Category 3. A summary of the category breakdown is given in Table 5-1.

**Table 5-1. Number of Subjects in Each SWAT Sorting Category.**

| Category | | Females | Males | Total |
|---|---|---|---|---|
| 1 | Good Sorts | 20 | 79 | 99 |
| 2 | Iterative Sorts | 1 | 4 | 5 |
| 3 | Poor Sorts | 7 | 12 | 19 |
| Total | | 28 | 95 | 123 |

## 5.2 Prototype Groups

Individual SWAT scales may be developed for each subject or group scales may be developed by averaging the ranks of the 27 descriptors across subjects. One advantage of producing group scales is that any errors subjects might have made in the SWAT sort tend to be averaged, thus providing a better approximation of the underlying construct. On the other hand, development of a collective scale for all subjects tends to obscure differences in weighting that may exist for specific individuals (Reid et al., 1982).

One approach that maximizes grouping strength while minimizing the loss due to obscuring individual differences is to form homogeneous subgroupings. This is accomplished in SWAT by correlating individual rankings with "SWAT model prototype" rankings. The prototypes are based upon the assumptions that the three dimensions of Time, Effort and Stress (T, E and S) are combined according to an additive rule and that the prototype rankings represent perfect data with consistent weightings assigned to each of the three dimensions and no axiom violations.

The six possible model prototypes are TES, TSE, ETS, EST, SET, and STE -- where TES indicates the highest weighting is placed on Time and lowest weighting on Stress. In the card sort, this is indicated by values on the stress dimension changing most rapidly and values on the time dimension changing least rapidly. A subject's ranking may be correlated separately with the six prototype rankings. The highest correlation(s) identifies the dimension(s) to which the subject attributes thr highest weight and the prototype that best describes the subject's ratings. A correlation of 1.0 indicates that the subject performed an iterative sort corresponding to a specific prototype ranking (a Category 2 sorter above).

The SWAT analysis program computes a Spearman rank correlation between each subject's sort and the model sorts that would result from a perfect iterative ordering of the cards basec on :ne six prototypes. These correlations were used to assign subjects to one of four prototype groups (Time, Effort, Stress and None). Each prototype group contained subjects with high correlations with the perfect sorts on that dimension. Subjects who identified equally with two or more dimensions were placed in the "None" category.

Although all subjects were assigned to a group, some subjects' data were not included in developing the SWAT scale solution for that group (Table 5-2). The nineteen (19) subjects in Category 3 were not included in the SWAT scale development phase due to the poor quality of their sorts. In addition, fourteen (14) subjects were not included in the prototype group solutions due to a lack of positive identification with a particular prototype group.

In summary, twenty-one (21) female subjects and eighty-three (83) male subjects were included in two separate "whole group" solutions. Twenty (20) females and seventy (70) males were assigned to the Time, Effort or Stress groups for the prototype solutions. Separate prototype solutions were developed for females, males and the overall group.

### 5.3 SWAT Scaling Solutions

Eleven different SWAT scaling solutions were generated. A single solution including all subjects could not be developed due to program restrictions in the number of allowable subjects per run. For each solution, Kendall's coefficient of concordance was examined (Table 5-3). This coefficient provides an indication of how well the subjects in a particular group agreed in their card ordering during the sorting phase of the SWAT.

**Table 5-2. Number (Row Percentage) of Subjects in Each Prototype Group.**

| Subset | Prototype Group | | | | | | | | Total |
|---|---|---|---|---|---|---|---|---|---|
| | Time | | Effort | | Stress | | None | | |
| **Prototype Solutions** | | | | | | | | | |
| Females | 3 | (15) | 2 | (10) | 15 | (75) | | | 20 |
| Males | 17 | (24) | 10 | (14) | 43 | (62) | | | 70 |
| Total | 20 | (22) | 12 | (13) | 58 | (65) | | | 90 |
| **Whole Group Solution** | | | | | | | | | |
| Females | 3 | (14) | 2 | (10) | 15 | (71) | 1 | (5) | 21 |
| Males | 17 | (21) | 11 | (13) | 44 | (53) | 11 | (13) | 83 |
| **Subjects in Study** | | | | | | | | | |
| Females | 4 | (14) | 2 | (7) | 20 | (72) | 2 | (7) | 28 |
| Males | 18 | (19) | 14 | (15) | 52 | (55) | 11 | (11) | 95 |
| Total | 22 | (18) | 16 | (13) | 72 | (58) | 13 | (11) | 123 |

Table 5-3. **Kendall Coefficients of Concordance.**

| Subjects | Prototype Groups Time | Effort | Stress | Whole Group |
|---|---|---|---|---|
| Females | 0.893 | 0.979 | 0.866 | 0.778 |
| Males | 0.890 | 0.827 | 0.850 | 0.727 |
| All | 0.884 | 0.836 | 0.851 | |

The coefficients for the "whole group" solutions (0.78,0.73) agree with the findings of Reid et al. (1982), where the coefficient was 0.76 before dividing the subjects into the three prototype groups. Kendall's coefficients of concordance for the prototype groups are also comparable to those found by Reid (0.83-0.98 vs. 0.96 for the Effort group and 0.85-0.87 vs. 0.90 for the Stress group). The Reid study did not have any subjects in the Time group. The Kendall's coefficients of concordance in the current study for the Time group ranged from 0.88 to 0.89.

From these data and other studies performed by the authors, it is believed that larger group sizes (males vs. females) result in reduced homogeneity of the subjects, leading to lower Kendall's coefficients. While this study used 20 female and 70 male subjects to obtain the solutions, the Reid study used seven subjects (two in the Stress group and five in the Effort group). The coefficients found in this study encourage the use of the prototype group solutions which yield greater agreement among subjects.

The SWAT scaling conversion may be performed by means of a conversion table or through the use of factors generated by the solution program. These scaling factors (identified on the SWAT program output as Timel, Time2, Time3, Effortl, Effort2, Effort3, Stressl, Stress2 and Stress3) may be selectively added to compute the scaled SWAT value. For example, a SWAT rating of 3-1-3 produces a scaled value as follows: SWAT = Time3 + Effortl + Stress3. A 1-1-1 rating and a 3-3-3 rating for Time, Effort and Stress define the adjusted scale end points of 0 and 100 respectively. As a result, Timel + Effortl + Stressl = 0 and Time3 + Effort3 + Stress3 = 100.

The resulting sort factors from the various scaling solutions for the current study are given in Tables 5-4 through 5-6 and the resulting SWAT scales are provided in Appendix B-1.

Table 5-4. Sort Factors - Female Subjects.

| Factor | Whole Group | Prototype Groups | | |
|---|---|---|---|---|
| | | Time | Effort | Stress |
| | *n*=21 | *n*=3 | *n*=2 | *n*=15 |
| Time1 | 2.54 | -13.89 | 6.54 | 4.06 |
| Time2 | 11.44 | 15.65 | 12.56 | 9.49 |
| Time3 | 28.19 | 44.62 | 16.77 | 24.21 |
| Effort1 | 0.93 | 4.39 | -5.24 | 2.82 |
| Effort2 | 18.12 | 18.46 | 6.02 | 19.86 |
| Effort3 | 31.69 | 28.34 | 45.28 | 29.80 |
| Stress1 | -3.47 | 9.50 | -1.29 | -6.89 |
| Stress2 | 17.88 | 16.76 | 13.91 | 19.87 |
| Stress3 | 40.12 | 27.04 | 37.94 | 45.99 |
| Time3 - Time1 | 25.65 | 58.51 | 10.24 | 20.15 |
| Effort3 - Effort1 | 30.75 | 23.95 | 50.53 | 26.97 |
| Stress3 - Stress1 | 43.60 | 17.54 | 39.23 | 52.88 |

Table 5-5. Sort Factors - Male Subjects.

| Factor | Whole Group | Prototype Groups | | |
|---|---|---|---|---|
| | | Time | Effort | Stress |
| | *n*=83 | *n*=17 | *n*=10 | *n*=43 |
| Time1 | 0.11 | -14.05 | 5.68 | 5.82 |
| Time2 | 11.29 | 11.94 | 14.15 | 11.91 |
| Time3 | 27.08 | 42.12 | 22.63 | 24.07 |
| Effort1 | 2.85 | 8.41 | -11.26 | 2.61 |
| Effort2 | 17.45 | 18.66 | 20.51 | 19.76 |
| Effort3 | 31.95 | 25.86 | 43.40 | 28.93 |
| Stress1 | -2.96 | 5.64 | 5.59 | -8.44 |
| Stress2 | 20.34 | 20.98 | 17.59 | 16.94 |
| Stress3 | 40.97 | 32.02 | 33.97 | 47.00 |
| Time3 - Time1 | 26.97 | 56.17 | 16.96 | 18.25 |
| Effort3 - Effort1 | 29.10 | 17.45 | 54.66 | 26.32 |
| Stress3 - Stress1 | 43.93 | 26.37 | 28.38 | 55.43 |

**Table 5-6. Sort Factors - All Subjects.**

| Factor | Prototype Groups | | |
|---|---|---|---|
| | Time | Effort | Stress |
| | *n* =20 | *n* =12 | *n* =58 |
| Time1 | -14.35 | 4.33 | 5.00 |
| Time2 | 13.28 | 14.63 | 12.00 |
| Time3 | 43.64 | 22.24 | 24.00 |
| Effort1 | 7.80 | -9.30 | 2.66 |
| Effort2 | 18.24 | 14.64 | 18.83 |
| Effort3 | 26.08 | 45.56 | 29.38 |
| Stress1 | 6.55 | 4.96 | -7.66 |
| Stress2 | 19.51 | 18.03 | 16.55 |
| Stress3 | 30.28 | 32.20 | 46.13 |
| Time3 - Time1 | 57.99 | 17.91 | 19.49 |
| Effort3 - Effort1 | 18.28 | 54.86 | 26.72 |
| Stress3 - Stress1 | 23.73 | 27.23 | 53.79 |

The differential sensitivity of the three prototypes is evident from the bottom sections of Tables 5-4 through 5-6. The magnitude of the difference between the highest and lowest values of the variables for each dimension (e.g., Time3 - Time1) gives an indication of the sensitivity of each prototype group to that dimension. These ranges averaged 22 scale points for the dimensions that differ from the prototype dimension but increased to approximately 58, 54 and 54 scale points for the Time, Effort and Stress dimensions matched with the Time, Effort and Stress prototypes respectively. This sensitivity was reduced in the "whole group" solution which emphasized the stress dimension due to the large number of subjects in that prototype.

Correlations were computed among the scale values produced by the "whole group" and the time, effort and stress solutions (Table 5-7). The results showed a high correlation between the "whole group" solutions and the prototype group solutions (0.77,0.83 with Time, 0.88,0.87 with Effort and 0.98,0.97 with Stress) with somewhat lower correlations among the prototype groups. This indicates that the "whole group" solution reflected the views of all subjects regardless of their prototype group. In general, the individual prototype groups did not possess as high a mutual agreement.

**Table 5-7. Correlation of Scale Values for the Whole Group Solution and the Prototype Group Solutions.**

| | | Prototype Groups | | |
|---|---|---|---|---|
| | Whole Group | Time | Effort | Stress |
| **Females** | | | | |
| Whole Group | 1.00 | .77 | .88 | .98 |
| Time | | 1.00 | .56 | .67 |
| Effort | | | 1.00 | .85 |
| Stress | | | | 1.00 |
| **Males** | | | | |
| Whole Group | 1.00 | .83 | .87 | .97 |
| Time | | 1.00 | .64 | .71 |
| Effort | | | 1.00 | .81 |
| Stress | | | | 1.00 |
| **Combined** | | | | |
| Time | | 1.00 | .64 | .70 |
| Effort | | | 1.00 | .80 |
| Stress | | | | 1.00 |

Although the prototype solutions provided a higher sensitivity for their respective dimensions, a rank ordering of the scale values for a given prototype solution (Appendix B-2) showed a substantial deviation from the perfect (iterative) ordering of the SWAT descriptors for that prototype as described in Reid et al. (1982).

The results of these separate analyses (Kendall's coefficient of concordance, magnitude of difference between highest and lowest sort factor of each dimension, and correlation of scale values) indicated that using prototype group solutions tends to improve the level of group concordance while maintaining the prototype preferences of each subject.

The selection of a particular scale (solution) for each subject was based on the number of subjects available for that solution. Due to the small sample size of females in each prototype solution and the similarities between the female and male solution scales, the ratings for T, E and S prototyped females were converted using the T, E and S solutions for females and males combined. Ratings for females not assigned to a prototype group were converted using the female "whole group" solution. The ratings from male subjects were converted using the T, E and S solutions generated solely from male subjects. Ratings for males not assigned to a prototype group were converted using the male "whole group" solution.

## 5.4 SWAT Ratings - Range Across Tasks

Since workload represents a composition of task difficulty, subject ability, subject effort and other factors, it is understandable that the ratings for a particular task-level will differ among subjects. Also, rating sensitivity is expected to vary across subjects. Furthermore, the range of workload conditions to which a subject has been exposed is likely to influence the scale end points (minimum and maximum ratings) used by each subject. An indication of these factors (subject variability, rating sensitivity, range of workload conditions) is provided by examining the variability (standard deviation, minimum vs. maximum) of the SWAT ratings across the various tasks for each subject.

Using the baseline data (Trials 6 and 8), variability across the 25 task-level combinations was computed for individual subjects and various subgroups (e.g., by gender or prototype). The mean, standard deviation, min and max values for each subject are provided in Appendix B-3 with a sample of three subjects and summaries for the gender and prototype subgroups given in Table 5-8.

Subject #7 provides an example of an individual who used the full range of SWAT ratings to evaluate the 25 tasks. This resulted in a high overall mean rating, a large standard deviation and a high maximum rating. Subject #10 represents the other extreme with a significantly lower maximum rating (25.4 for Trial 6), a much lower mean, and a much smaller standard deviation. Subject #16 falls between the two extremes. With few exceptions, subjects provided ratings of 0.0 for the lowest workload conditions ($\bar{x}_{MIN}$ = 1.2). Across subjects for Trials 6 and 8, the average of the maximum ratings was 74.5.

Differences between males and females and among the various prototypes were negligible for the standard deviation, min and max values. Differences in means for these subgroups will be examined in more detail in the next section. In summary, there were substantial differences in the range of ratings given by individual subjects to the 25 tasks. However, these differences in rating sensitivity do not appear related to gender or prototype differences.

**Table 5-8. SWAT Rating Variability Across Tasks.**
**(Summary Across 25 Task-Level Combinations)**

| | Trial 6 | | | | Trial 8 | | | |
|---|---|---|---|---|---|---|---|---|
| | Mean | Std | Min | Max | Mean | Std | Min | Max |
| **Subject ID** | | | | | | | | |
| #07 | 40.7 | 24.7 | 0.0 | 100.0 | 51.9 | 23.9 | 0.0 | 92.8 |
| #10 | 7.2 | 9.1 | 0.0 | 25.4 | 11.7 | 13.7 | 0.0 | 48.6 |
| #16 | 21.9 | 19.4 | 0.0 | 69.9 | 23.0 | 19.7 | 0.0 | 59.4 |
| $\bar{x}$ | 27.1 | 20.8 | 1.1 | 75.6 | 26.7 | 21.0 | 1.2 | 73.4 |
| **Gender** | | | | | | | | |
| Female (*n*=28) | 28.9 | 26.2 | 0.0 | 100.0 | 26.6 | 26.7 | 0.0 | 100.0 |
| Male (*n*=95) | 26.6 | 24.7 | 0.0 | 100.0 | 26.7 | 25.1 | 0.0 | 100.0 |
| **Prototype** | | | | | | | | |
| Time (*n*=22) | 26.5 | 26.2 | 0.0 | 100.0 | 26.5 | 27.0 | 0.0 | 100.0 |
| Effort (*n*=16) | 30.9 | 25.5 | 0.0 | 100.0 | 29.5 | 25.6 | 0.0 | 100.0 |
| Stress (*n*=72) | 26.9 | 24.5 | 0.0 | 100.0 | 26.9 | 25.2 | 0.0 | 100.0 |
| None (*n*=13) | 24.7 | 25.1 | 0.0 | 100.0 | 22.2 | 23.3 | 0.0 | 100.0 |

## 5.5 SWAT Ratings - Baseline Trials

The mean SWAT ratings by task and difficulty level are summarized in Table 5-9 for Trials 6 and 8 and presented in Figures 5-1 through 5-9. The *r* values in the figures represent the Trial 6-Trial 8 intercorrelations which ranged from 0.43 for Spatial Processing at the medium level to 0.80 for Interval Production (mean = 0.66, std = 0.09, median = 0.67). The ratings are further summarized by gender in Table 5-13 and Figures 5-13 through 5-21. Only slight differences between trials or between genders were observed in the data. Univariate summaries of the SWAT rating distributions for each task-level combination are presented in Appendix B-4. Data for these summaries included the Trial 6 and Trial 8 data for all 123 subjects.

Table 5-9. Mean SWAT Ratings by Task, Level and Trial.

| Level | Low | | Medium | | High | |
|---|---|---|---|---|---|---|
| Trial | 06 | 08 | 06 | 08 | 06 | 08 |
| Task | | | | | | |
| CR | 19.1 | 17.5 | 34.8 | 34.1 | 55.5 | 56.5 |
| GR | 26.1 | 28.5 | 36.7 | 36.3 | 50.8 | 49.4 |
| IP | 7.7 | 7.1 | . | . | . | . |
| LP | 11.9 | 11.4 | 22.2 | 23.4 | 26.8 | 28.3 |
| MP | 12.3 | 11.7 | 23.4 | 22.7 | 30.8 | 30.5 |
| MS | 7.9 | 7.4 | 13.5 | 16.2 | 26.5 | 26.8 |
| PM | 11.6 | 9.9 | 30.6 | 32.1 | 46.4 | 49.3 |
| SP | 9.0 | 5.3 | 19.9 | 17.0 | 27.3 | 22.3 |
| UT | 20.3 | 18.7 | 43.9 | 42.5 | 63.0 | 61.7 |

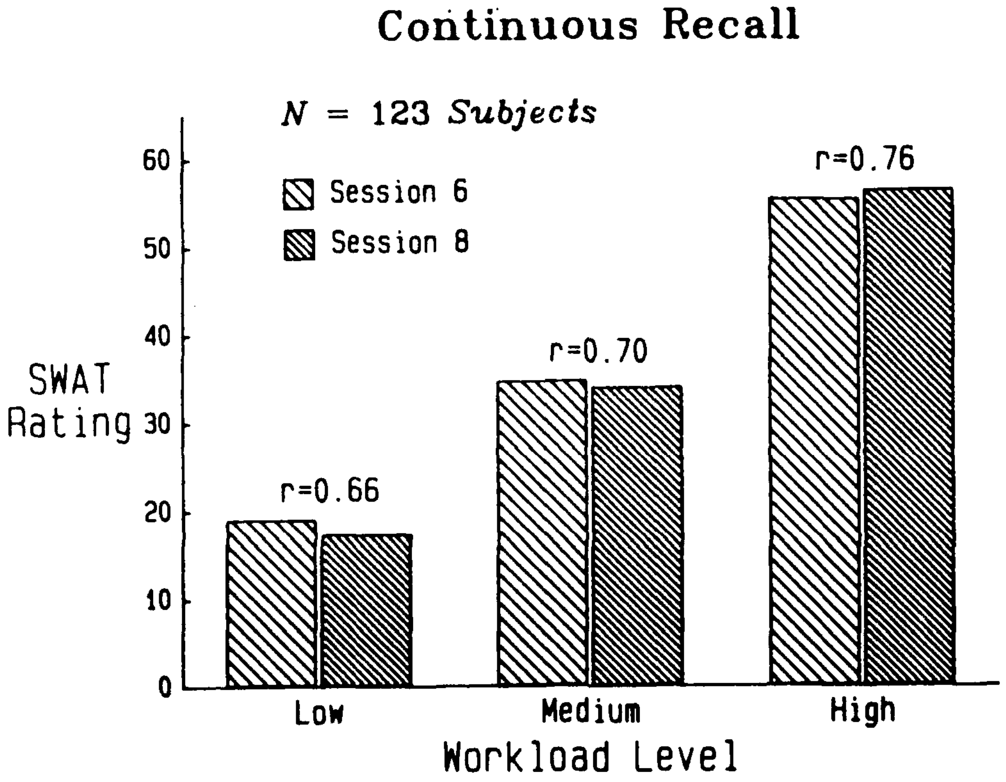


**Figure 5-1. Mean SWAT Ratings for Continuous Recall - Trials 6 and 8.**

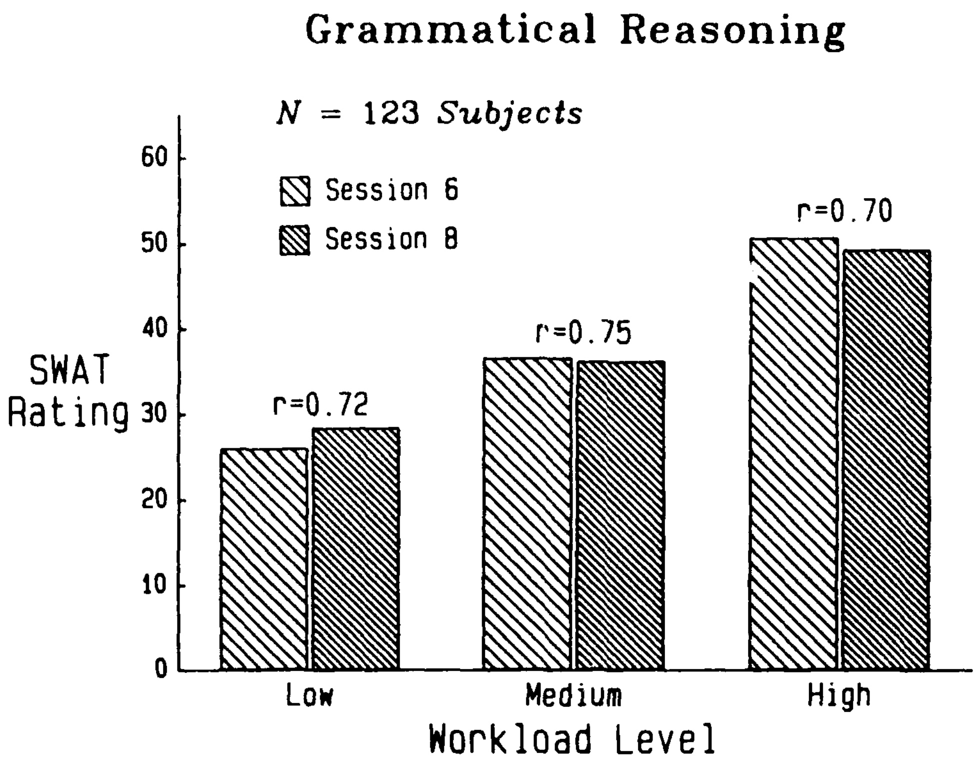


Figure 5-2. Mean SWAT Ratings for Grammatical Reasoning - Trials 6 and 8.

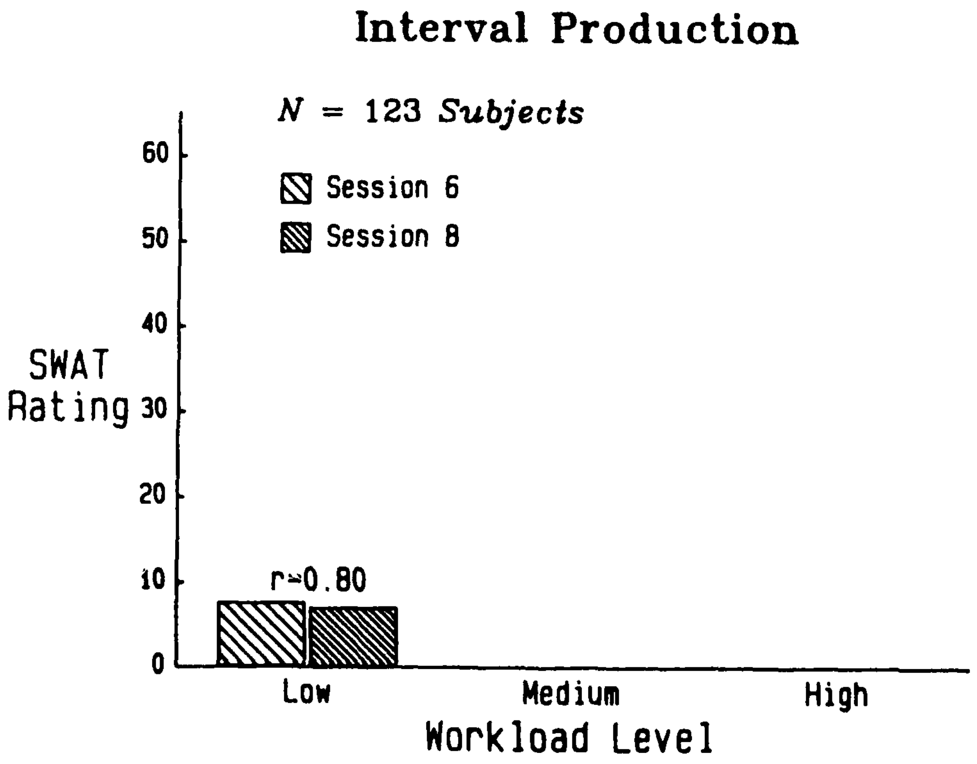


Figure 5-3. Mean SWAT Ratings for Interval Production - Trials 6 and 8.

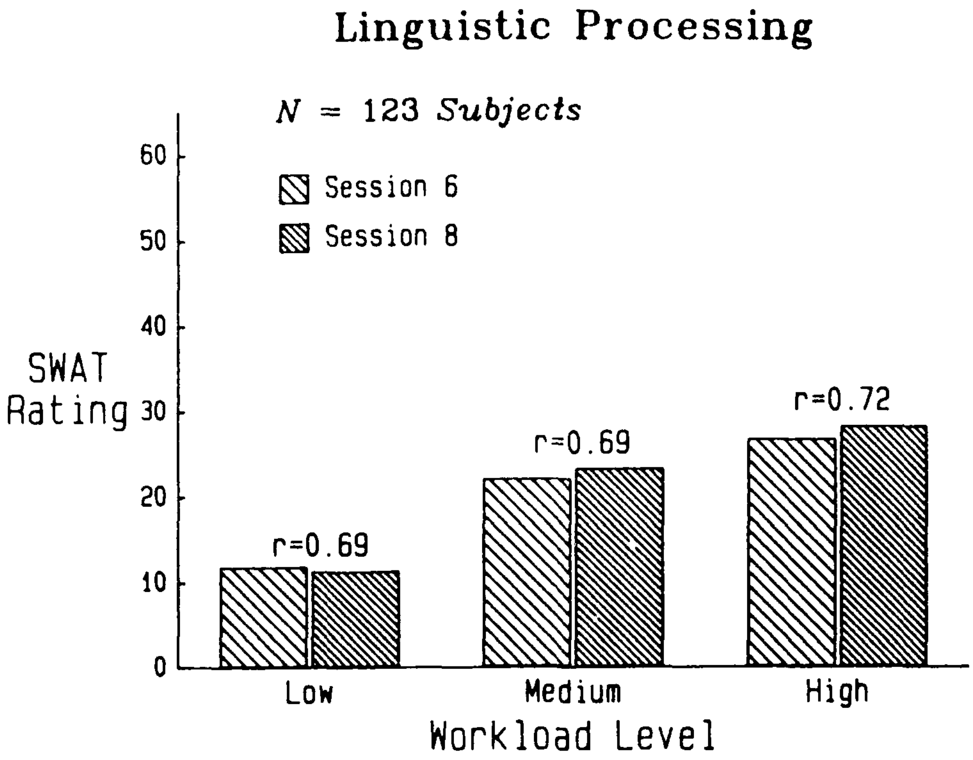


Figure 5-4. Mean SWAT Ratings for Linguistic Processing - Trials 6 and 8.

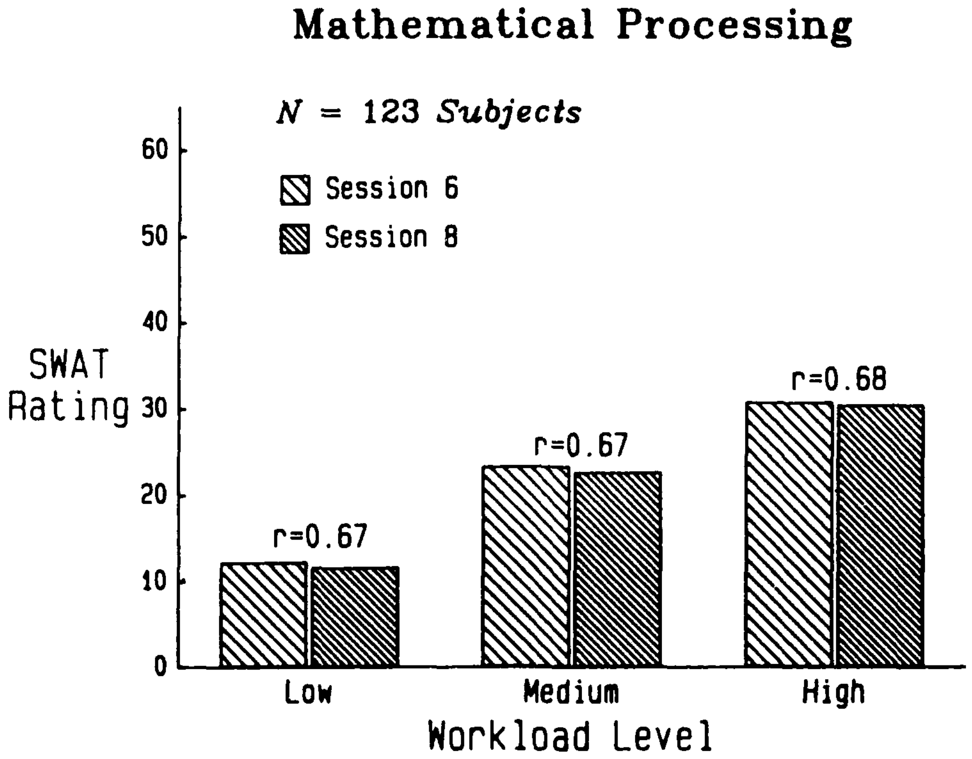


Figure 5-5. Mean SWAT Ratings for Mathematical Processing - Trials 6 and 8.

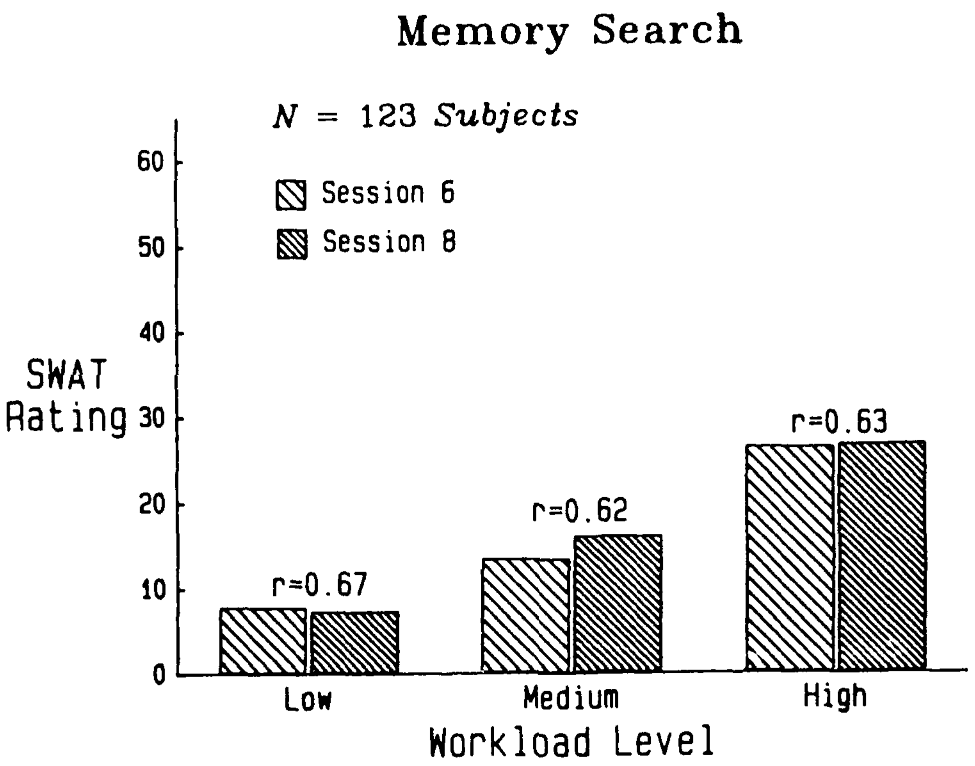


**Figure 5-6. Mean SWAT Ratings for Memory Search - Trials 6 and 8.**

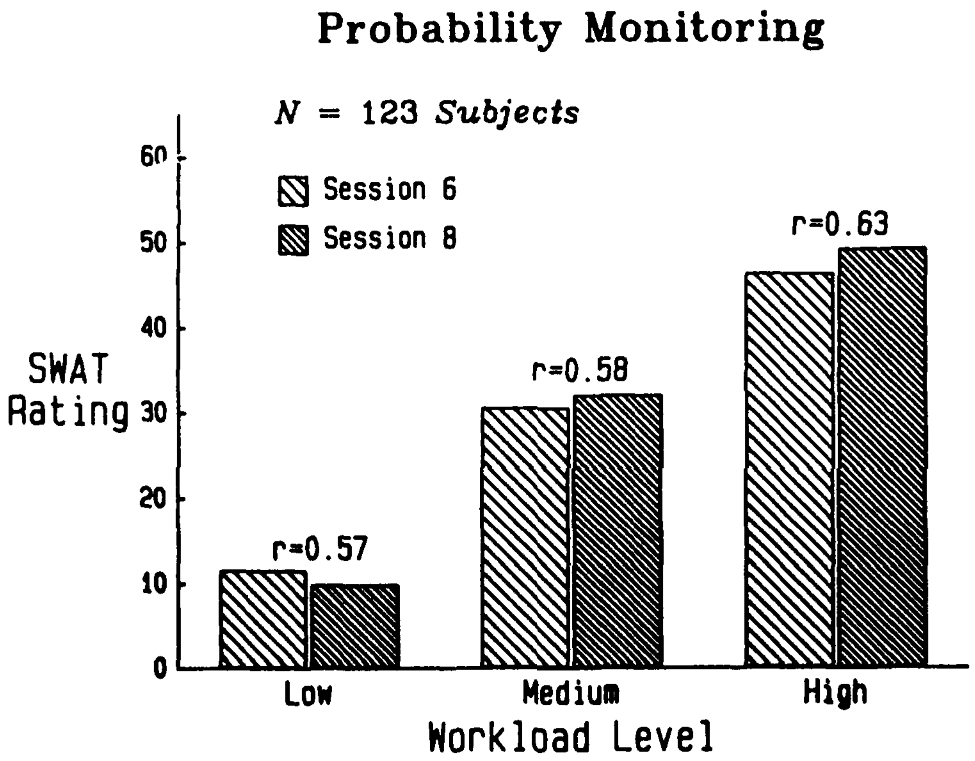


**Figure 5-7. Mean SWAT Ratings for Probability Monitoring - Trials 6 and 8.**

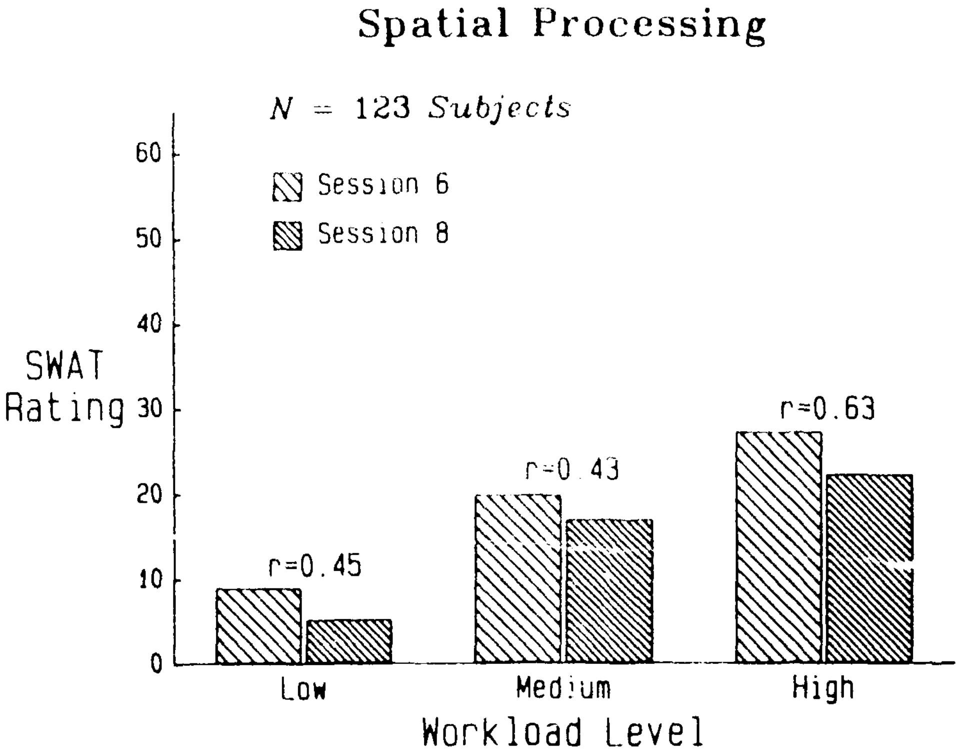


Figure 5-8. Mean SWAT Ratings for Spatial Processing - Trials 6 and 8.

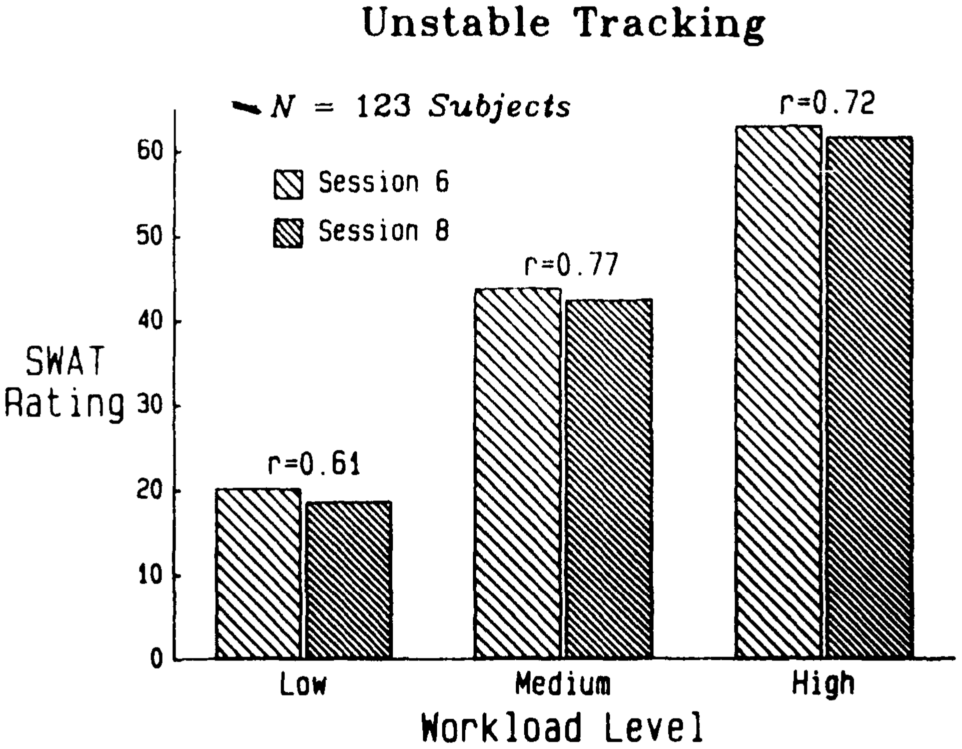


Figure 5-9. Mean SWAT Ratings for Unstable Tracking - Trials 6 and 8.

### 5.5.1 Analysis of Task and Trial Differences

Various models were used to investigate differences in SWAT ratings between tasks and between trials. The major differences between models involved the treatment of the various task-level combinations and whether subject variability and subject interactions were included in the model. For all models, task levels differed substantially while there was no significant difference between Trial 6 and Trial 8 SWAT values. Subject variability was substantial and the subject interactions with task and trial were significant.

The model which provided the highest $R^2$ value (0.91) was a repeated measures design in which the various task-level combinations were treated as a single factor with 25 levels as follows:

$$SWAT_{ijk} = \mu + L_i + T_j + LT_{ij} + S_k + LS_{ik} + TS_{jk} + \varepsilon_{ijk}$$

where:

$L_i$ = Task-Level combination, e.g., CR-LOW, $i = 1, ..., 25$
$T_j$ = Trial (6 vs. 8), $j = 1, 2$
$S_k$ = Subject, $k = 1, ..., 123$.

Seven observations were missing out of 6150. The results of this analysis are summarized in Table 5-10. Note that for this model, **Subject** is treated as a random factor and the appropriate denominators in the $F$-ratios for level and trial are their respective interactions with subject.

**Table 5-10. ANOVA Summary for Task and Trial Effects.**

| Source | DF | Mean Square | F | p > F | % Variance |
|---|---|---|---|---|---|
| Task-Level (L) | 24 | 59109 | 164.52 | 0.0001 | 36 |
| Trial (T) | 1 | 330 | 0.62 | 0.4344 | 0 |
| L by T | 24 | 220 | 1.85 | 0.0073 | 0 |
| Subject (S) | 122 | 8445 | 70.90 | 0.0001 | 26 |
| L by S | 2928 | 359 | 3.02 | 0.0001 | 18 |
| T by S | 122 | 536 | 4.50 | 0.0001 | 3 |
| Error | 2921 | 119 | | | 18 |

The significant differences between tasks reflect differences in difficulty levels and differences between the tasks themselves. To verify the sensitivity of the SWAT to task difficulty manipulations, a separate analysis was performed for each task using the following model:

$$SWAT_{ijk} = \mu + L_i + T_j + LT_{ij} + S_k + LS_{ik} + TS_{jk} + \varepsilon_{ijk}$$

where:

$L_i$ = Level, $i = 1, 2, 3$<br>
$T_j$ = Trial (6 vs. 8), $j = 1, 2$<br>
$S_k$ = Subject, $k = 1, ..., 123$.

The results of the analysis are summarized in Table 5-11. A Tukey studentized range test demonstrated that *all three difficulty levels differed significantly for all tasks.* In only one instance (SP) was there a significant difference between the Trial 6 and Trial 8 SWAT ratings with Trial 8 having the lower average. There were no significant level by trial interactions.

**Table 5-11. ANOVA Summary for Level and Trial Effects by Task.**

| Task | Model | $F_{Level}$ | SWAT Mean | | | $F_{Trial}$ |
|---|---|---|---|---|---|---|
| | $R^2$ | (2,244) | Low | Med | High | (1,122) |
| CR | 0.95 | 318.11* | 18.3 | 34.4 | 56.0 | 0.26 |
| GR | 0.95 | 154.00* | 27.3 | 36.5 | 50.1 | 0.88 |
| IP | 1.00 | . | 7.4 | . | . | 0.40 |
| LP | 0.94 | 90.62* | 11.7 | 22.8 | 27.5 | 0.57 |
| MP | 0.94 | 152.15* | 12.0 | 23.0 | 30.6 | 0.23 |
| MS | 0.93 | 153.19* | 7.7 | 14.8 | 26.7 | 0.64 |
| PM | 0.94 | 297.81* | 10.8 | 31.3 | 47.9 | 0.53 |
| SP | 0.93 | 116.34* | 7.2 | 18.5 | 24.8 | 9.46** |
| UT | 0.95 | 271.05* | 19.5 | 43.2 | 62.4 | 1.44 |

* $p<0.0001$

** $p=0.0026$ (Trial 8 significantly lower than Trial 6)

### 5.5.2 Magnitude of Level Effect for Time, Effort and Stress Dimensions

In addition to statistical significance, the magnitude of the difficulty manipulation effect on the SWAT ratings was estimated using a procedure described by Vaughan and Corballis (1969) and applied by Vidulich and Tsang (1986). The procedure involves calculating variance estimates for the factors in an analysis of variance and expressing these estimates as the percentage of total variance accounted for by each factor. The percentage of variance accounted for provides an indication of the relative sensitivity of different dependent measures to the difficulty manipulations. The absolute variance estimates allow comparison between experiments or between different dependent variables employing the same units of measurement.

In this application, the variance estimates provide a combined indication of (1) the relative magnitude of the difficulty manipulation, and (2) the sensitivity of the SWAT ratings to this manipulation, across the eight CTS tasks that exist at three workload levels. Similar variance estimates employing the unscaled Time, Effort and Stress ratings indicate the relative sensitivity of each dimension to the difficulty manipulation. A comparison of the percentage of variance accounted for across the three dimensions indicates the relative loading on each dimension for each task. Estimates were also computed for the subject effect to indicate relative subject variability across tasks and across the Time, Effort and Stress dimensions.

The computational procedures of Dodd and Schultz (1973) were used to obtain the estimates. In agreement with the ANOVA model used in this study, the Dodd and Schultz formulas for a fixed factor, repeated-measures design with a partially additive model (no Level by Trial by Subject interaction) were used (Dodd and Schultz, Table 2, p. 392). Estimates for the overall model are presented in Table 5-10 while estimates for each task are presented in Table 5-12 and in Figures 5-10 through 5-12. The Range score in Table 5-12 represents the difference in ratings between the high and low levels of the task.

The breakdown in Table 5-10 for the overall model shows that 35% of the total variability in the SWAT ratings was attributable to actual differences in task difficulty, while 26% was a result of subject variability. Thus, the magnitude of subject variability was approximately 75% of the magnitude of the task effect. An additional 18% of the variability was attributable to the task-by-subject interaction with approximately 18% remaining as random error.

Table 5-12. Magnitude of Level and Subject Effects by Task.

| Task | Measure | Total Variance | Difficulty Level | | | Subject | |
|---|---|---|---|---|---|---|---|
| | | | $F_{(2,244)}$ | % Var. | Range | $F_{(122,244)}$ | % Var. |
| CR | SWAT | 686.46 | 318.11 | 35 | 37.7 | 16.33 | 35 |
| | Time | 0.67 | 66.04 | 5 | 0.5 | 31.50 | 70 |
| | Effort | 0.42 | 308.56 | 43 | 1.1 | 3.91 | 15 |
| | Stress | 0.45 | 145.92 | 24 | 0.8 | 10.57 | 34 |
| GR | SWAT | 666.28 | 154.00 | 13 | 22.8 | 23.92 | 56 |
| | Time | 0.63 | 51.66 | 4 | 0.4 | 33.60 | 71 |
| | Effort | 0.34 | 106.76 | 13 | 0.5 | 10.78 | 41 |
| | Stress | 0.46 | 85.98 | 10 | 0.5 | 13.21 | 48 |
| LP | SWAT | 423.23 | 90.62 | 10 | 15.8 | 19.13 | 52 |
| | Time | 0.56 | 7.80 | 0 | 0.1 | 49.17 | 80 |
| | Effort | 0.34 | 98.67 | 18 | 0.6 | 5.75 | 30 |
| | Stress | 0.27 | 18.48 | 3 | 0.3 | 10.36 | 43 |
| MP | SWAT | 445.56 | 152.15 | 13 | 18.6 | 19.55 | 53 |
| | Time | 0.56 | 17.75 | 1 | 0.1 | 47.97 | 83 |
| | Effort | 0.34 | 173.62 | 22 | 0.6 | 8.52 | 35 |
| | Stress | 0.27 | 33.22 | 4 | 0.2 | 9.17 | 37 |
| MS | SWAT | 408.13 | 153.19 | 14 | 19.0 | 15.94 | 48 |
| | Time | 0.54 | 26.30 | 2 | 0.3 | 65.21 | 79 |
| | Effort | 0.30 | 169.74 | 27 | 0.7 | 4.78 | 21 |
| | Stress | 0.19 | 37.93 | 6 | 0.2 | 7.06 | 38 |
| PM | SWAT | 728.72 | 297.81 | 31 | 37.1 | 11.91 | 33 |
| | Time | 0.69 | 85.62 | 9 | 0.6 | 19.56 | 55 |
| | Effort | 0.44 | 278.13 | 36 | 0.9 | 5.86 | 23 |
| | Stress | 0.46 | 120.33 | 17 | 0.7 | 10.69 | 37 |
| SP | SWAT | 411.19 | 116.34 | 13 | 17.6 | 14.15 | 39 |
| | Time | 0.38 | 24.65 | 2 | 0.3 | 36.21 | 64 |
| | Effort | 0.34 | 173.68 | 23 | 0.6 | 6.91 | 29 |
| | Stress | 0.21 | 23.75 | 3 | 0.2 | 8.14 | 34 |
| UT | SWAT | 1005.91 | 271.05 | 30 | 42.9 | 18.12 | 37 |
| | Time | 0.81 | 86.70 | 9 | 0.6 | 31.08 | 62 |
| | Effort | 0.59 | 220.65 | 27 | 0.9 | 10.8 | 34 |
| | Stress | 0.62 | 164.61 | 22 | 1.0 | 14.00 | 38 |

For individual tasks, the total variance exhibited a strong direct relationship with overall task difficulty. This was also true of the absolute variance estimates for both the level and subject factors. In other words, higher variability was associated with the more difficult tasks.

Across individual tasks there was general agreement from the various measures of level effect magnitude (*F*-ratio, absolute variance estimate, percentage of variance accounted for, and range). The percentage of variance accounted for by the level effect ranged from 10% for LP to 35% for CR (Figure 5-10). In addition, there was a direct relationship between overall task difficulty (as measured by the ratings) and magnitude of effect. That is, tasks with higher overall workload ratings (CR,UT) in general demonstrated a greater level effect. This is also evident from the Range measure where a higher range indicates a greater difficulty spread based on the subjective ratings.

For all tasks, subject variability exceeded the level effect with ratios ranging from 1:1 for CR to 5:1 for LP. The percentage of variance accounted for by subject variability ranged from 33% for PM to 56% for GR. Zero variability was attributed to the trial factor.

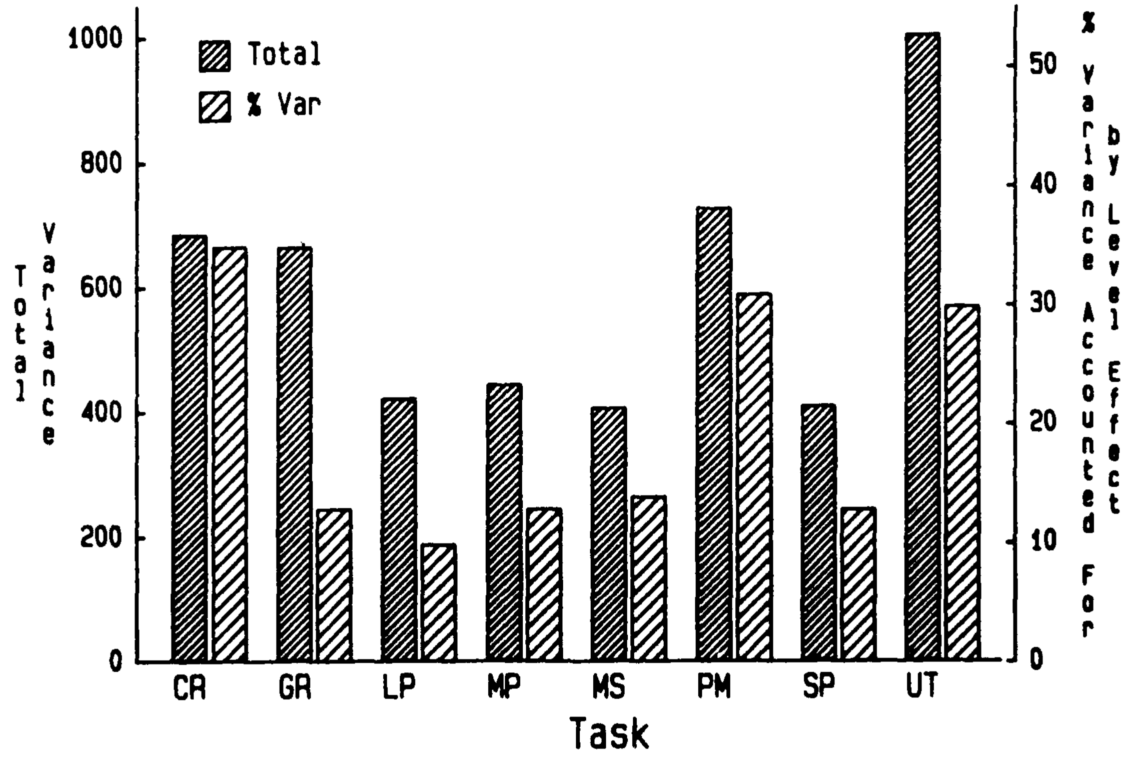


**Figure 5-10. Total Variance and Percent Accounted for by Level Factor.**

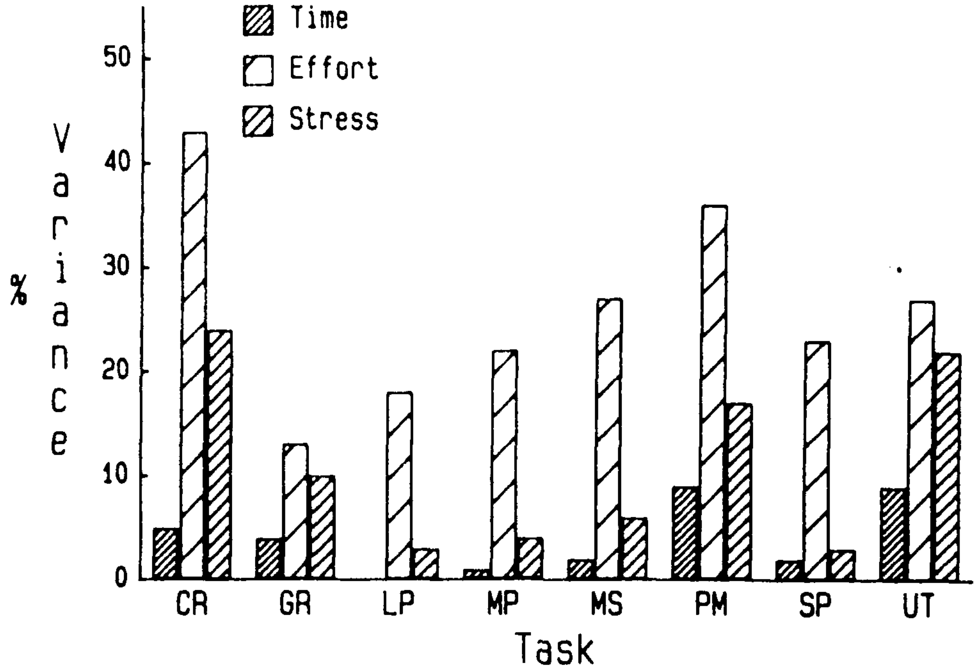


Figure 5-11. Percent Variance for Time, Effort, Stress - Level Effect.

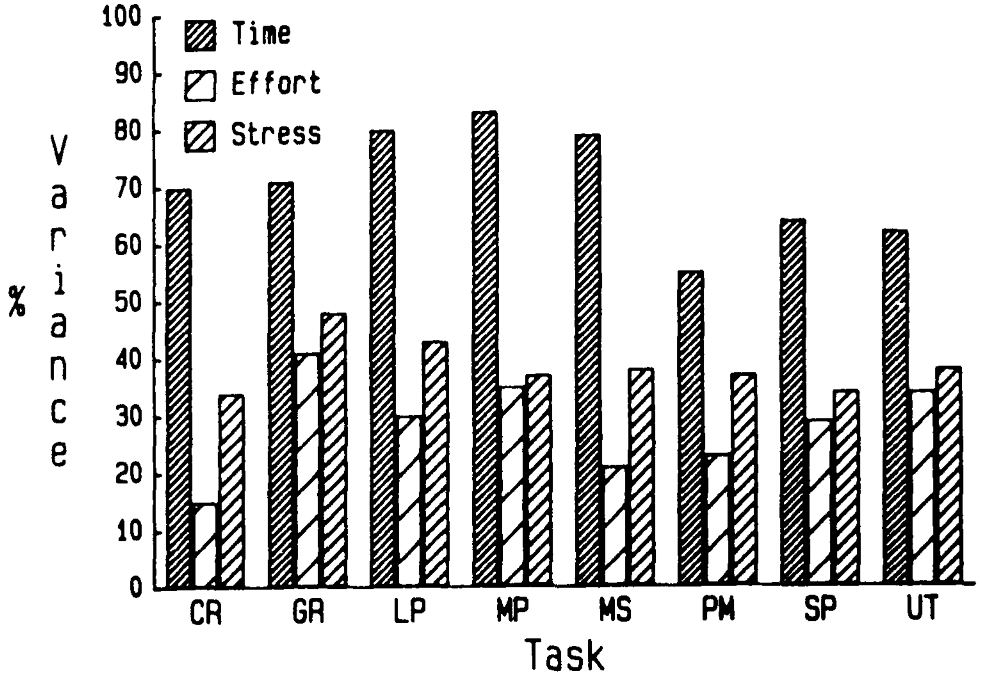


Figure 5-12. Percent Variance for Time, Effort, Stress - Subject Effect.

Separate analyses of the Time, Effort and Stress ratings showed similar patterns with respect to the order of magnitudes across tasks (Figure 5-11). For all eight tasks, the Effort rating was most sensitive to changes in task difficulty level (13% to 43%), followed by Stress (3% to 24%) and Time (0% to 9%). The magnitudes of the level effect for the Time ratings are only notable (9%) for the Probability Monitoring and Unstable Tracking tasks. In terms of the construct validity of the SWAT dimensions, this is important since PM and UT are the only tasks that are machine paced and require continuous attention to the display (higher time demands). These two tasks also produced nontrivial accountable variability on the Stress dimension as did Continuous Recall and Grammatical Reasoning.

Absolute subject variability was highest and fairly constant along the Time dimension, averaging 0.42, followed by 0.14 for Stress and 0.11 for Effort. The percentage of variance accounted for by subjects averaged 71% for Time, 29% for Effort and 39% for Stress (Figure 5-12). In general, there was an inverse relationship between the magnitude of the level effect and the amount of subject variability. Subjects were highly variable in their ratings on the Time dimension and these ratings were only slightly related to task difficulty. On the other hand, the subject variability was low on the Effort dimension but the effect of task difficulty was large implying that subjects uniformly employed this dimension to distinguish between the task levels. The Stress dimension fell between the other two with a moderate amount of subject variability and moderate sensitivity to task difficulty.

These results are further verified by Tukey tests ($\alpha = 0.01$) performed on the Time, Effort and Stress ratings. The Effort ratings were able to distinguish between all three levels for all eight tasks. The Time and Stress ratings had this ability for only six of the eight tasks.

Table 5-13. Mean SWAT Ratings by Task, Level and Gender for Trials 6 and 8 Combined.

| Level | Low | | Med | | High | |
|---|---|---|---|---|---|---|
| Gender | Fem | Male | Fem | Male | Fem | Male |
| Task | | | | | | |
| CR | 15.6 | 19.1 | 31.0 | 35.4 | 52.1 | 57.1 |
| GR | 27.4 | 27.3 | 37.0 | 36.4 | 48.9 | 50.5 |
| IP | 6.1 | 7.8 | . | . | . | . |
| LP | 12.3 | 11.5 | 22.8 | 22.8 | 29.5 | 26.9 |
| MP | 12.8 | 11.8 | 22.9 | 23.1 | 29.4 | 31.0 |
| MS | 6.1 | 8.1 | 13.2 | 15.3 | 25.1 | 27.1 |
| PM | 14.9 | 9.5 | 35.8 | 30.0 | 53.4 | 46.2 |
| SP | 11.7 | 5.8 | 21.4 | 17.6 | 26.2 | 24.4 |
| UT | 24.2 | 18.1 | 49.2 | 41.5 | 64.0 | 61.9 |

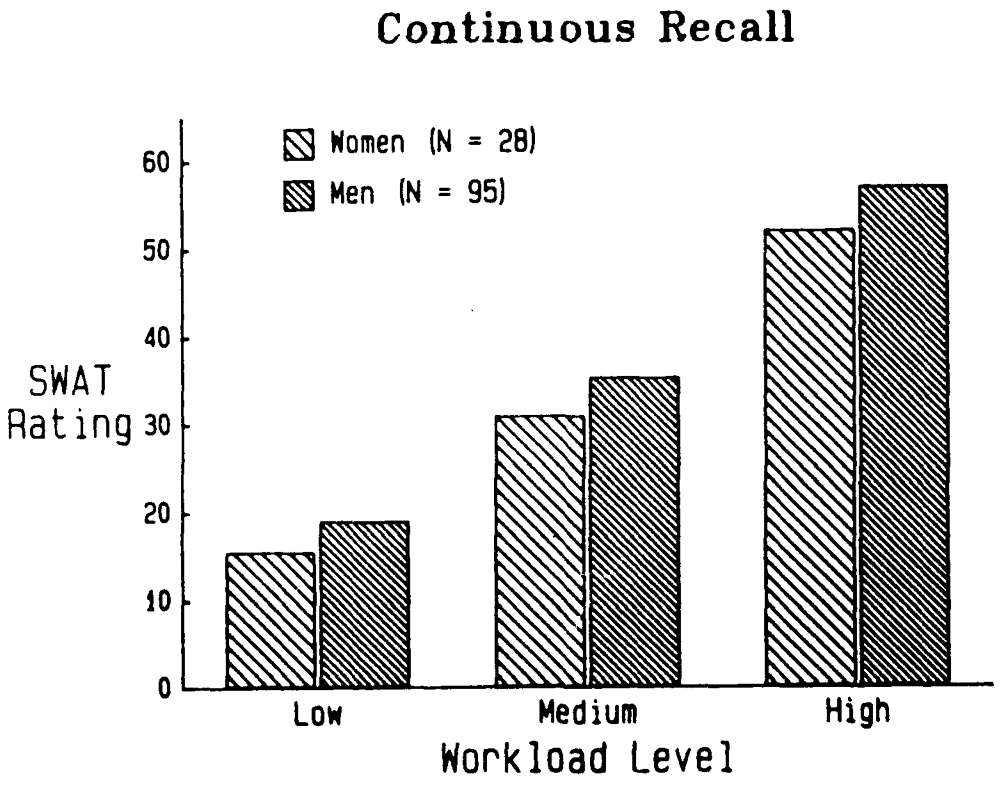


Figure 5-13. SWAT Ratings for Continuous Recall - Men vs. Women.

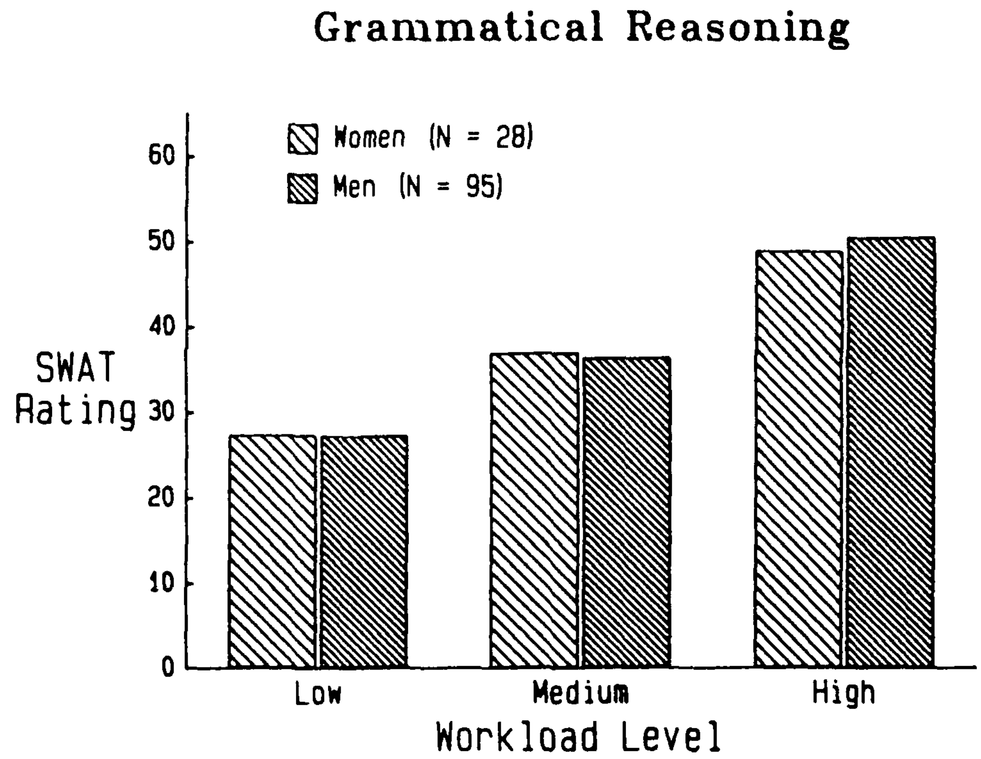


Figure 5-14. SWAT Ratings for Grammatical Reasoning - Men vs. Women.

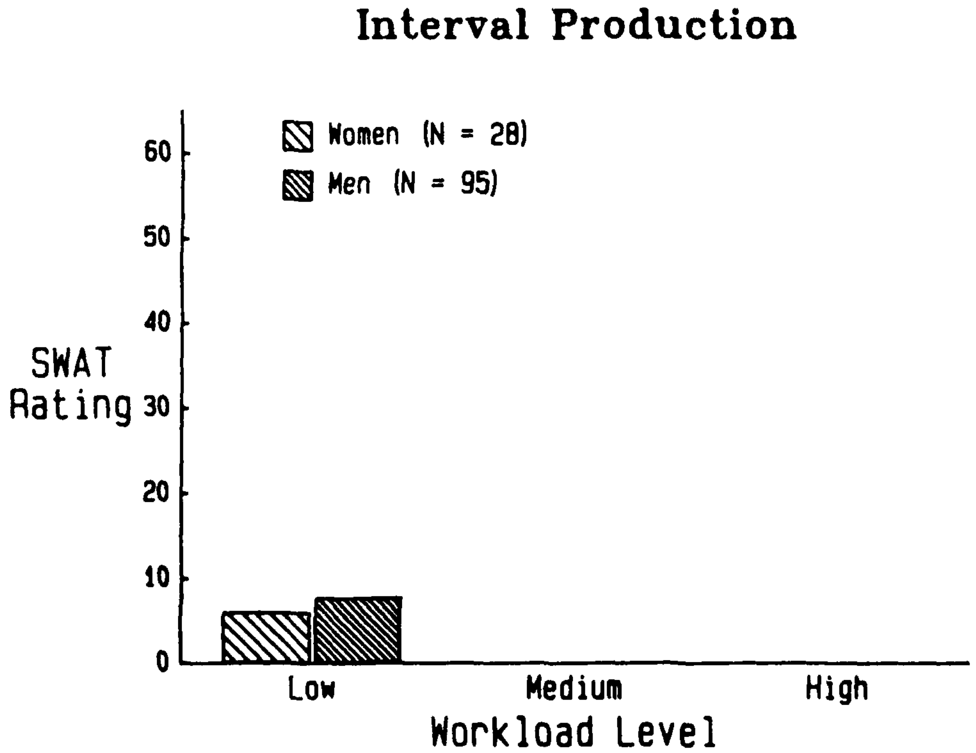


Figure 5-15. SWAT Ratings for Interval Production - Men vs. Women.

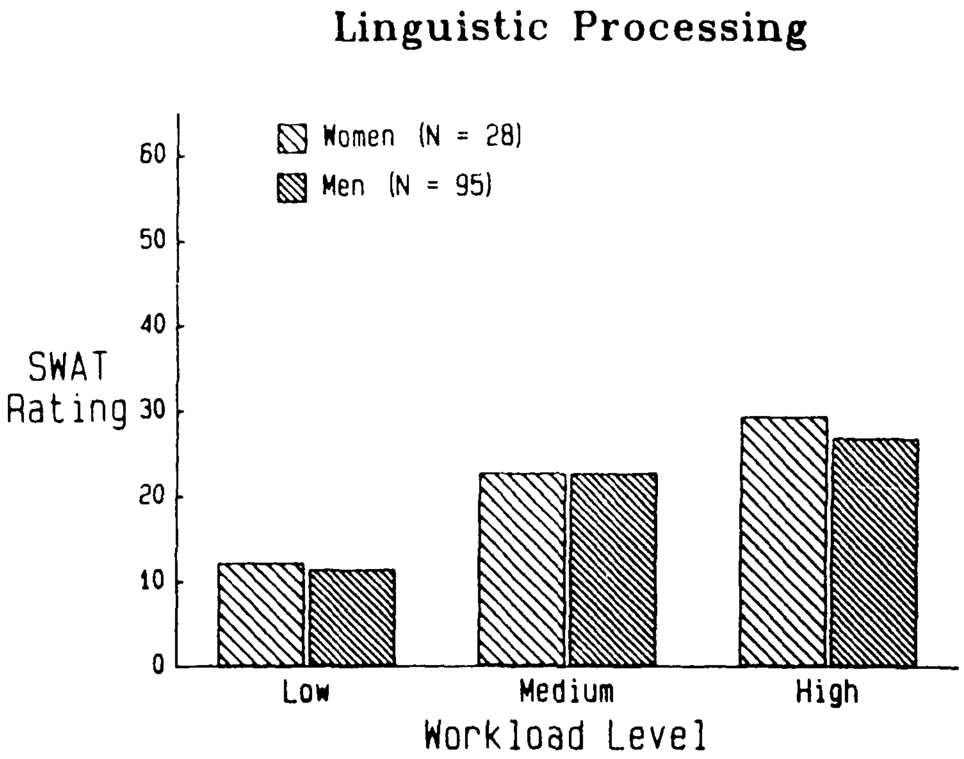


Figure 5-16. SWAT Ratings for Linguistic Processing - Men vs. Women.

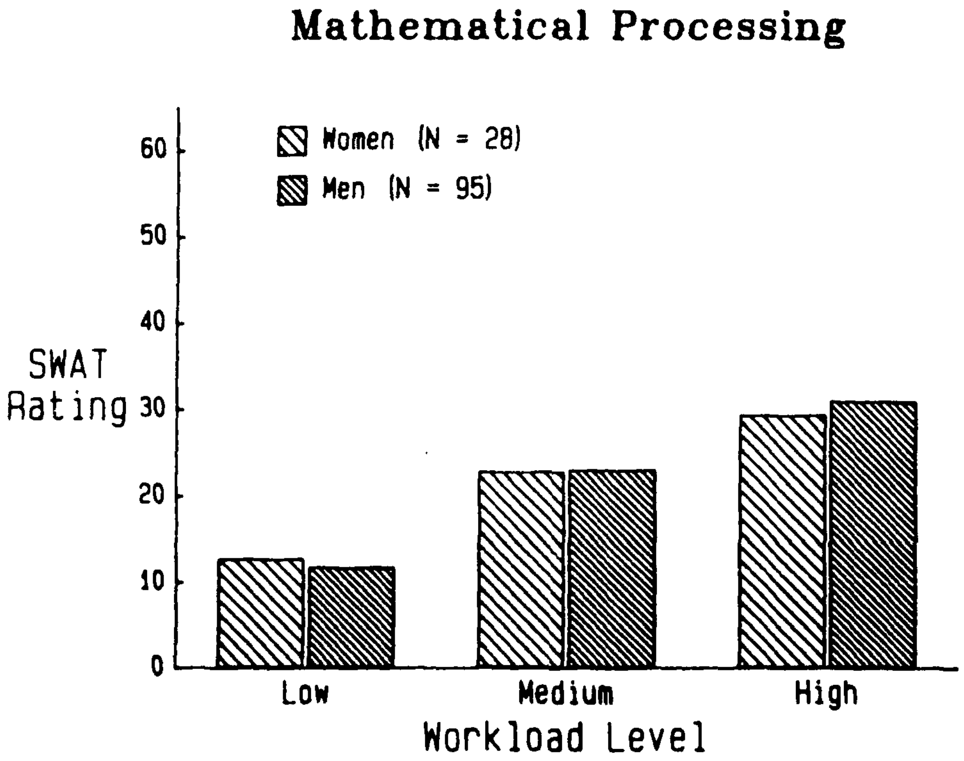


Figure 5-17. SWAT Ratings for Mathematical Processing - Men vs. Women.

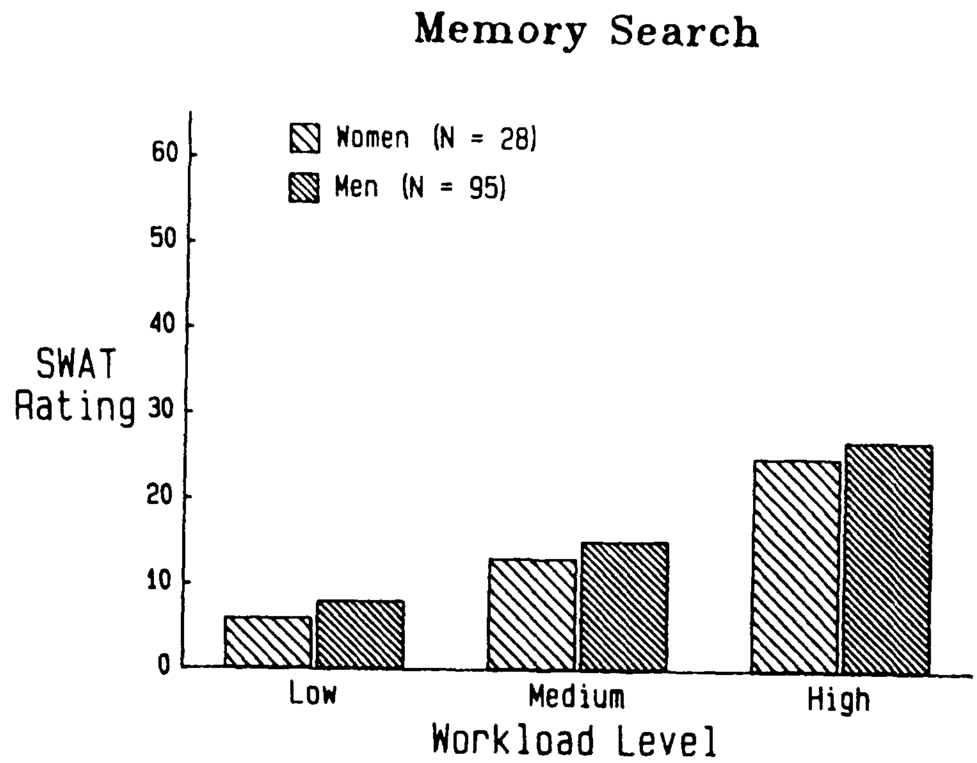


Figure 5-18. SWAT Ratings for Memory Search - Men vs. Women.

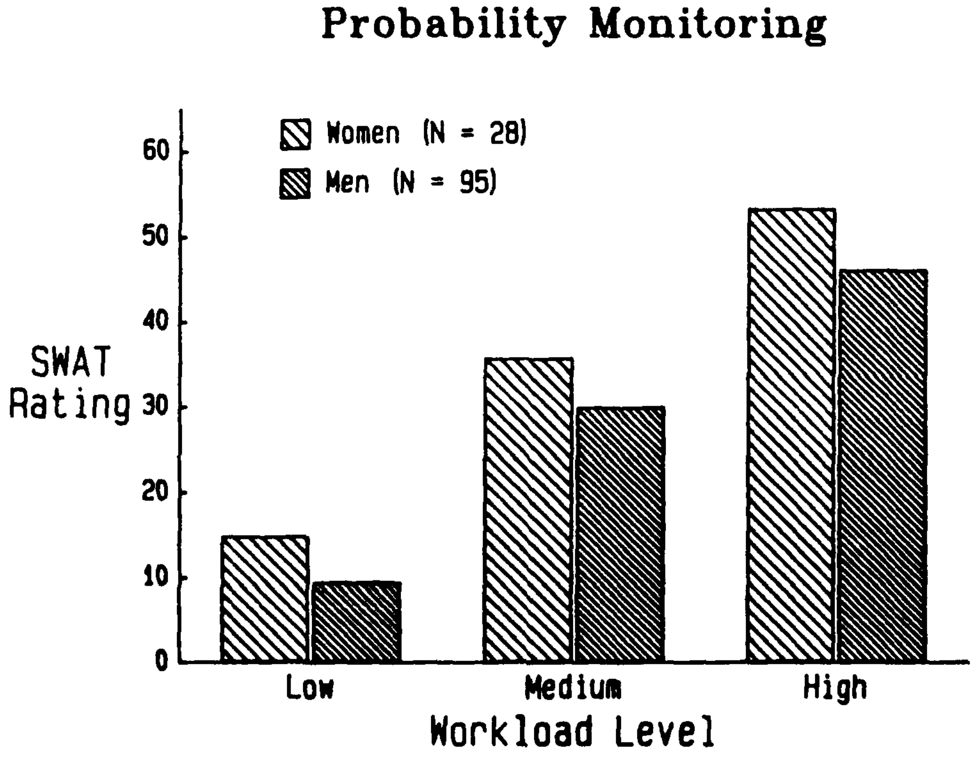


Figure 5-19. SWAT Ratings for Probability Monitoring - Men vs. Women.

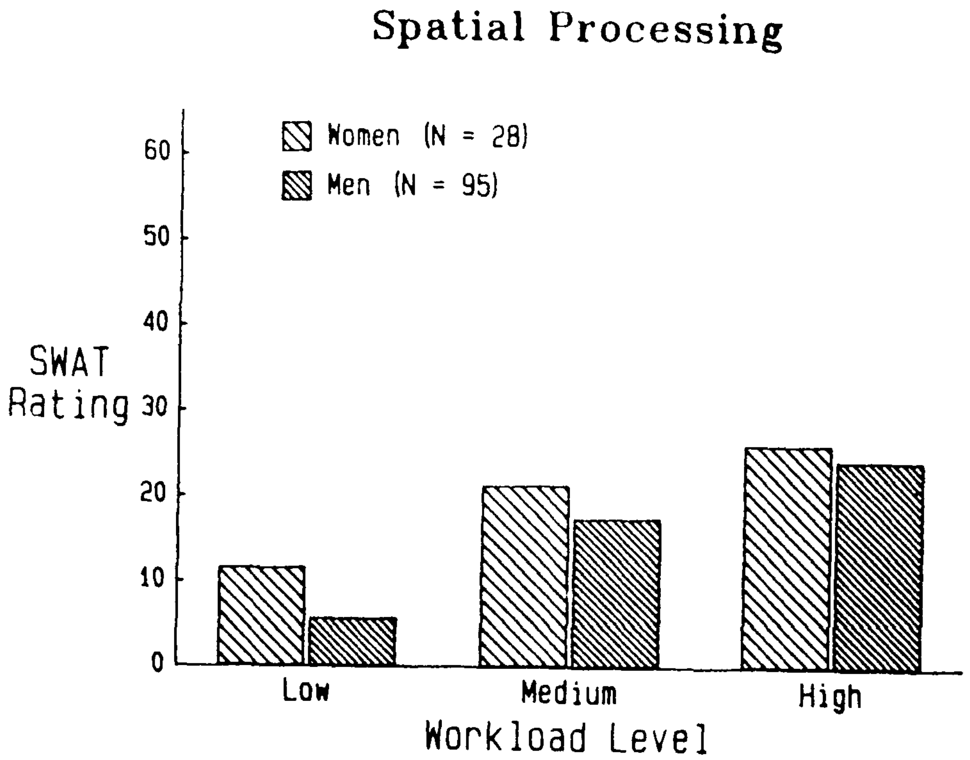


Figure 5-20. SWAT Ratings for Spatial Processing - Men vs. Women.

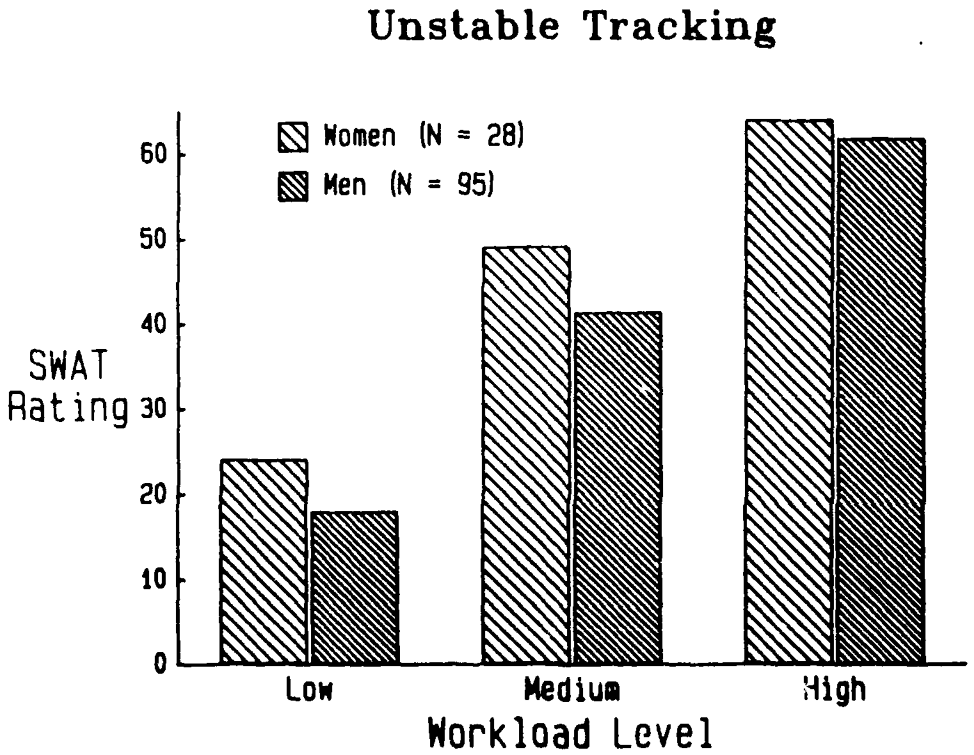


Figure 5-21. SWAT Ratings for Unstable Tracking - Men vs. Women.

### 5.5.3 Analysis of Gender and Prototype Differences

To identify possible gender or prototype rating differences, additional analyses were performed by further partitioning of the subject variability. The previously mentioned ANOVA models (Section 5.5.1) were used with the addition of the factor GENDER (or PROTOTYPE) and its interactions with Task-Level (or Level) and with Trial. The Subject factor thus becomes nested within the grouping variable (gender or prototype). The results of these analyses are summarized in Tables 5-14 through 5-16.

**Table 5-14. ANOVA Summary for Gender Effect.**

| Source | DF | Sum of Squares | Mean Square | F | p > F |
|---|---|---|---|---|---|
| Task-Level (L) | 24 | 1418607 | 59109 | 165.38 | 0.0001 |
| Trial (T) | 1 | 330 | 330 | 0.62 | 0.4311 |
| L by T | 24 | 5279 | 220 | 1.85 | 0.0073 |
| Gender (G) | 1 | 1245 | 1245 | 0.15 | 0.7027 |
| L by G | 24 | 14087 | 587 | 1.64 | 0.0254 |
| T by G | 1 | 1464 | 1464 | 2.77 | 0.0987 |
| Subject (S) | 121 | 1029061 | 8505 | 71.40 | 0.0001 |
| L by S | 2904 | 1037904 | 357 | 3.00 | 0.0001 |
| T by S | 121 | 63977 | 529 | 4.44 | 0.0001 |
| Error | 2921 | 347937 | 119 | | |

**Table 5-15. ANOVA Summary for Prototype Effect.**

| Source | DF | Sum of Squares | Mean Square | F | p > F |
|---|---|---|---|---|---|
| Task-Level (L) | 24 | 1418607 | 59109 | 165.04 | 0.0001 |
| Trial (T) | 1 | 330 | 330 | 0.61 | 0.4363 |
| L by T | 24 | 5279 | 220 | 1.85 | 0.0073 |
| Prototype (P) | 3 | 16701 | 5576 | 0.65 | 0.5822 |
| L by P | 72 | 29125 | 405 | 1.13 | 0.2152 |
| T by P | 3 | 1075 | 358 | 0.66 | 0.5768 |
| Subject (S) | 119 | 1013604 | 8518 | 71.51 | 0.0001 |
| L by S | 2856 | 1022866 | 358 | 3.01 | 0.0001 |
| T by S | 119 | 64366 | 541 | 4.54 | 0.0001 |
| Error | 2921 | 347937 | 119 | | |

Table 5-14 indicates that there was no significant difference between men and women across the 25 task-level combinations. However, there was a marginal task-level by gender interaction indicating possible gender differences for specific tasks. Table 5-15 shows that the ratings for the tasks did not differ among the various prototype classifications. Table 5-16 summarizes the analyses for the individual tasks.

Although the differences were not statistically significant, females gave lower ratings than males to those tasks with a high memory component (CR, MS) and gave higher ratings than males to those tasks involving input/output and spatial elements (PM, SP, UT). Among the prototype groups, the Effort group gave consistently higher ratings across most tasks while subjects with no prototype classification gave lower ratings. Again, these differences were not statistically significant.

**Table 5-16. ANOVA Summary for Gender and Prototype Effects by Task.**

| Task | $F_{Gender}$ | SWAT | | $F_{Ptype}$ | SWAT Mean | | | |
|---|---|---|---|---|---|---|---|---|
| | (1,121) | Fem | Male | (3,119) | Time | Eff | Str | None |
| CR | 1.58 | 32.9 | 37.2 | 0.80 | 33.1 | 40.3 | 36.8 | 33.2 |
| GR | 0.00 | 37.8 | 38.0 | 0.27 | 36.9 | 41.7 | 37.9 | 35.6 |
| IP | 0.32 | 6.1 | 7.8 | 0.29 | 8.7 | 5.4 | 7.1 | 9.4 |
| LP | 0.12 | 21.6 | 20.4 | 0.13 | 20.1 | 20.9 | 21.2 | 18.5 |
| MP | 0.00 | 21.7 | 21.9 | 0.65 | 20.2 | 26.5 | 21.8 | 19.4 |
| MS | 0.42 | 14.8 | 16.8 | 0.19 | 15.9 | 16.5 | 17.0 | 13.7 |
| PM | 3.06* | 34.7 | 28.6 | 3.33+ | 31.4 | 39.7 | 28.9 | 21.8 |
| SP | 1.87 | 19.8 | 15.9 | 2.22† | 15.1 | 22.6 | 17.1 | 10.5 |
| UT | 1.54 | 45.8 | 40.5 | 0.27 | 45.0 | 41.8 | 41.1 | 39.7 |
| $\bar{x}$ | | 27.7 | 26.7 | | 26.5 | 30.2 | 26.9 | 23.5 |

* p=0.08 + p=0.02 † p=0.09

## 5.6 Comparison with Previous Data

Table 5-17 presents a comparison of the current data with the data reported by Shingledecker (1984). The current data represents a summary (mean and std) of trials 6 and 8 for all 123 subjects.

In the current study, the ratings were lower for tasks CR, LP, MS, PM, and SP, higher for GR and MP and mixed (higher at one level, lower at another) for UT. These results were perhaps due to the fact that subjects in the current study trained on all CTS tasks concurrently while the parametric studies (Shingledecker, 1984) were performed in isolation. Thus, subjects in the current study may have been exposed to a wider range of workload conditions.

**Table 5-17. Comparison of Current Data with Shingledecker (1984).**

| Level | Low | | | Medium | | | High | | |
|---|---|---|---|---|---|---|---|---|---|
| Task | Current | | 1984 | Current | | 1984 | Current | | 1984 |
| | Mean | Std | Mean | Mean | Std | Mean | Mean | Std | Mean |
| CR | 18.3 | 17.3 | 23 | 34.4 | 18.4 | 39 | 56.0 | 24.2 | 67 |
| GR | 27.3 | 20.5 | 21 | 36.5 | 22.5 | 37 | 50.1 | 25.9 | 43 |
| IP | 7.4 | 14.2 | ? | - | - | - | - | - | - |
| LP | 11.7 | 16.0 | 23 | 22.8 | 19.5 | 33 | 27.5 | 20.0 | 47 |
| MP | 12.0 | 16.0 | 6 | 23.0 | 18.9 | 15 | 30.6 | 21.2 | 33 |
| MS | 7.7 | 14.4 | 23 | 14.8 | 17.2 | 35 | 26.7 | 21.1 | 59 |
| PM | 10.8 | 17.1 | 12 | 31.3 | 20.1 | 35 | 47.9 | 25.7 | 60 |
| SP | 7.2 | 14.7 | 10 | 18.4 | 17.6 | 18 | 24.8 | 19.9 | 33 |
| UT | 19.5 | 20.2 | 9 | 43.2 | 25.7 | 42 | 62.4 | 28.4 | 84 |
| $\bar{x}$ | 13.5 | 16.7 | 16 | 28.1 | 20.0 | 32 | 40.8 | 23.3 | 53 |
| Over All Task-Levels - $\bar{x}_{mean}$ = 26.9 $\bar{x}_s$ = 19.9 $\bar{x}_{1984}$ = 33 | | | | | | | | | |

## 5.7 Intertask Relationships

Table 5-18 presents the CTS tasks in ascending order of SWAT ratings for each difficulty level. Tasks with equivalent rankings (Tukey test, $\alpha = 0.05$) are shown within the same box.

Table 5-18. Subjective Ranking of Task Difficulty by Workload Level.

| Level | | | | | | Overall | |
|---|---|---|---|---|---|---|---|
| Low | | Medium | | High | | | |
| Task | SWAT | Task | SWAT | Task | SWAT | Task | SWAT |
| SP | 7.2 | - | - | - | - | IP | 7.4 |
| IP | 7.4 | MS | 14.8 | SP | 24.8 | MS | 16.4 |
| MS | 7.7 | SP | 18.4 | MS | 26.7 | SP | 16.8 |
| PM | 10.8 | LP | 22.8 | LP | 27.5 | LP | 20.7 |
| LP | 11.7 | MP | 23.0 | MP | 30.6 | MP | 21.9 |
| MP | 12.0 | PM | 31.3 | PM | 47.9 | PM | 30.0 |
| CR | 18.3 | CR | 34.4 | GR | 50.1 | CR | 36.2 |
| UT | 19.5 | GR | 36.5 | CR | 56.0 | GR | 38.0 |
| GR | 27.3 | UT | 43.2 | UT | 62.4 | UT | 41.7 |

### 5.7.1 Cluster Analysis

Cluster analysis of the SWAT ratings for the 25 task-level combinations produced the results given in Table 5-19. Four clusters were identified with clusters generally differentiated along the dimensions of task difficulty and processing stage. It is evident from the table that Cluster 1 contains discrete stimulus central processing tasks of low and moderate difficulty while Cluster 2 contains the more difficult central processing tasks. Cluster 3 contains the easy levels of the motor output tasks, a spatial task and a math processing task. Cluster 4 contains the difficult levels of the motor output task.

The results of the cluster analysis for the SWAT ratings help validate the Criterion Task Set as a battery of tasks that tap separate information processing resources and stages. However, some differences existed between the cluster structure for the performance data and that for the SWAT ratings indicating that subjects perform differently than their estimate of task difficulty as might be expected.

**Table 5-19. Cluster Analysis of SWAT Ratings.**

| Cluster | | | |
|---|---|---|---|
| 1 | 2 | 3 | 4 |
| CR1 | CR2 | UT1 | UT2 |
| | CR3 | IP1 | UT3 |
| GR1 | GR2 | | |
| | GR3 | | |
| MP2 | | MP1 | |
| MP3 | | | |
| LP1 | PM2 | PM1 | |
| LP2 | PM3 | | |
| LP3 | | | |
| MS1 | | SP1 | |
| MS2 | SP3 | SP2 | |
| MS3 | | | |

1 - low level 2 - medium level 3 - high level

## 5.8 SWAT Ratings - Training Trials

The means and standard deviations of the SWAT ratings for training trials 1 through 5 are presented by task and difficulty level in Table 5-20. The means are plotted in Figures 5-22 through 5-30. As seen in the figures, all tasks demonstrated a significant decline in the SWAT ratings over time, particularly from the first to the second trial. Many tasks continued to show a decrease in subjective difficulty through all five trials as subjects became more proficient and confident in their abilities. Analysis of variance was used to determine significance between trials for each of the tasks using the model presented in Section 5.5.1 on page 5-16. For those tasks with a significant trial by level interaction, separate analyses were performed for each level using a reduced model involving only the trial and subject effects. A summary of the ANOVA results from the first set of analyses is presented in Table 5-21. The individual task analyses and the results of Tukey studentized range tests are summarized in Table 5 22.

For the training trials, the SWAT ratings for all three difficulty levels differed significantly for all tasks except Linguistic Processing. SWAT ratings for the medium and high levels of the LP task did not differ and were in fact reversed for the first two trials, with the medium difficulty level having the highest rating. For all tasks and levels, there were no significant differences between trials 4 and 5, although a slight decreasing trend continued for eight of the nine tasks. When combined with the baseline trials, there were no significant differences among trials 4, 5, 6 and 8 for any of the tasks.

An investigation of possible gender or prototype differences during the training trials was conducted using the ANOVA model in Section 5.5.2. Ratings differed significantly between genders only for the Memory Search task ($F_{(1,121)} = 4.87$, $p < 0.029$) with $\bar{x}_{Men} = 20.9$ and $\bar{x}_{Women} = 15.5$. Rating trends on other tasks were similar to the baseline data with females giving lower ratings than males on the non-spatial central processing tasks and higher ratings than males on PM, SP and UT. These differences were particularly evident in the early trials. The trial by gender interaction was significant only for Mathematical Processing ($F_{(4,484)} = 3.02$, $p < 0.018$) indicating possible differences between genders for the trial-to-trial rating changes. At $\alpha = 0.01$, ratings did not differ among prototypes for any CTS task. A marginal difference ($F_{(3,119)} = 2.82$, $p < 0.042$) was observed for Probability Monitoring.

Table 5-20. Means (Standard Deviations) of SWAT Ratings by Task, Level and Trial - Training Trials.

| Task | Level | Trial | | | | |
|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 |
| CR | L | 41.8 (20.4) | 28.2 (18.6) | 22.1 (16.7) | 21.5 (18.7) | 19.8 (18.9) |
| | M | 56.9 (18.9) | 46.5 (18.5) | 38.5 (17.2) | 37.1 (18.2) | 34.1 (17.9) |
| | H | 67.9 (21.5) | 65.0 (21.1) | 63.1 (22.3) | 56.5 (22.9) | 58.3 (25.1) |
| GR | L | 39.0 (21.1) | 34.7 (19.8) | 30.9 (19.5) | 28.2 (19.1) | 27.4 (19.5) |
| | M | 41.0 (17.9) | 39.7 (18.2) | 37.8 (20.6) | 37.5 (20.2) | 37.0 (21.6) |
| | H | 55.4 (22.6) | 58.2 (25.1) | 53.7 (25.8) | 53.8 (26.0) | 50.7 (25.8) |
| IP | - | 14.7 (18.4) | 8.5 (12.4) | 8.7 (13.5) | 8.2 (13.5) | 7.9 (13.5) |
| LP | L | 13.3 (15.6) | 13.6 (16.9) | 11.9 (15.7) | 12.2 (16.7) | 11.2 (18.0) |
| | M | 30.4 (20.6) | 29.3 (19.1) | 25.3 (16.8) | 24.7 (18.4) | 22.5 (19.8) |
| | H | 28.9 (16.6) | 28.1 (17.9) | 26.0 (17.8) | 27.8 (19.2) | 26.2 (19.3) |
| MP | L | 17.1 (18.3) | 14.3 (16.7) | 13.4 (16.9) | 11.3 (15.3) | 12.0 (15.8) |
| | M | 30.1 (18.2) | 26.5 (17.0) | 25.3 (17.4) | 23.6 (18.2) | 23.2 (17.8) |
| | H | 34.7 (19.1) | 31.6 (18.5) | 32.5 (19.9) | 29.4 (19.4) | 30.6 (19.7) |
| MS | L | 19.7 (19.2) | 8.2 (12.8) | 6.3 (11.8) | 6.8 (11.9) | 6.5 (13.2) |
| | M | 28.2 (20.5) | 17.2 (15.8) | 15.4 (14.3) | 14.6 (16.5) | 15.7 (18.9) |
| | H | 43.1 (23.8) | 30.4 (18.5) | 29.0 (16.6) | 28.4 (18.8) | 25.8 (20.5) |
| PM | L | 21.0 (20.5) | 18.4 (19.8) | 15.2 (18.2) | 12.9 (16.8) | 12.9 (17.6) |
| | M | 48.5 (22.6) | 38.9 (22.6) | 33.9 (23.3) | 32.8 (21.9) | 31.1 (23.0) |
| | H | 60.7 (23.9) | 54.3 (26.0) | 51.4 (25.5) | 48.1 (25.8) | 47.2 (26.7) |
| SP | L | 7.4 (12.1) | 9.4 (14.2) | 6.5 (12.9) | 7.6 (13.6) | 6.3 (12.5) |
| | M | 24.4 (17.4) | 25.1 (17.0) | 20.2 (16.8) | 18.1 (15.5) | 19.4 (18.3) |
| | H | 32.1 (18.0) | 33.9 (19.6) | 30.1 (20.7) | 27.0 (19.5) | 25.2 (19.2) |
| UT | L | 37.8 (26.4) | 30.3 (25.2) | 27.1 (22.2) | 23.1 (22.5) | 18.6 (19.8) |
| | M | 64.0 (25.1) | 57.8 (25.1) | 52.9 (25.3) | 46.8 (26.5) | 43.8 (24.6) |
| | H | 69.8 (25.4) | 71.7 (25.9) | 70.8 (26.9) | 67.2 (27.8) | 64.1 (27.9) |

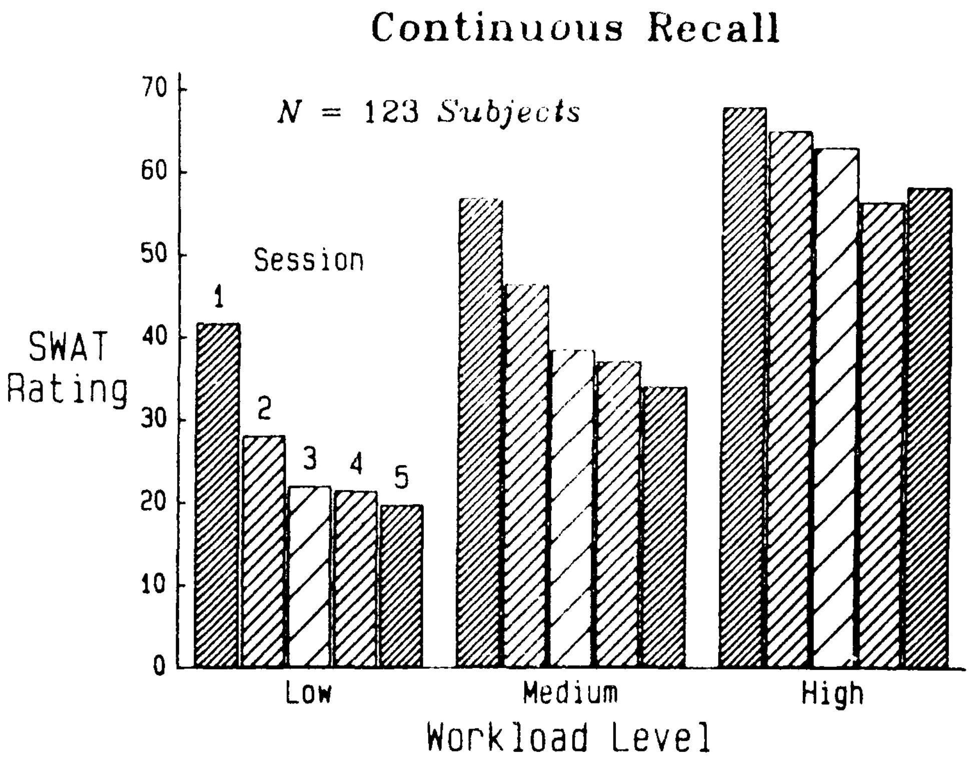


Figure 5-22. Mean SWAT Ratings for Continuous Recall - Trials 1 through 5.

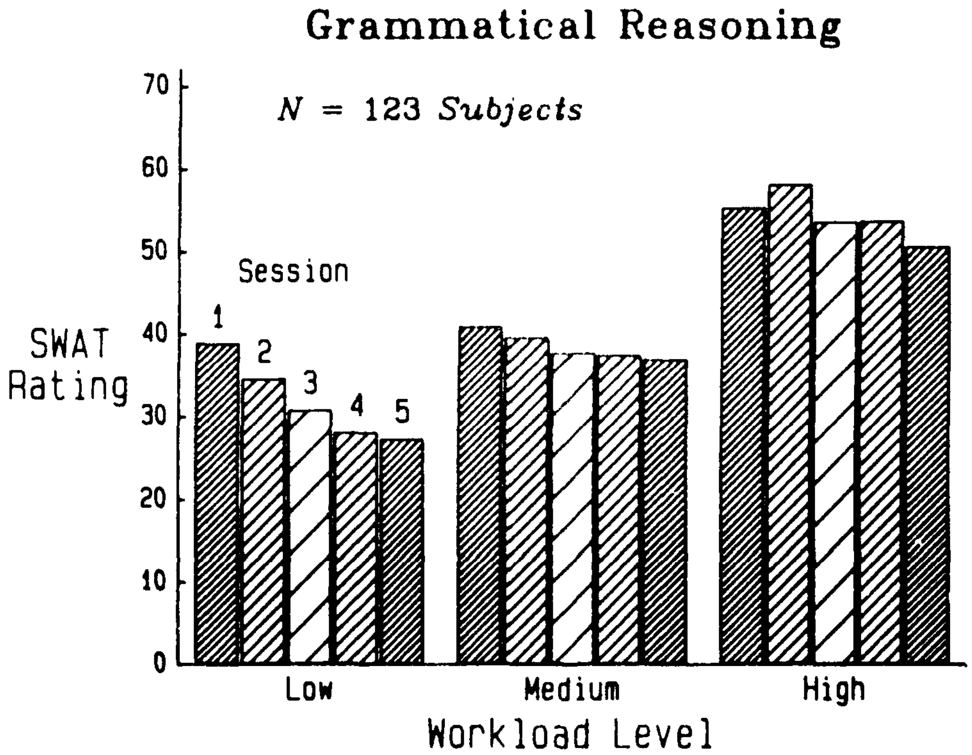


Figure 5-23. Mean SWAT Ratings for Grammatical Reasoning - Trials 1 through 5.

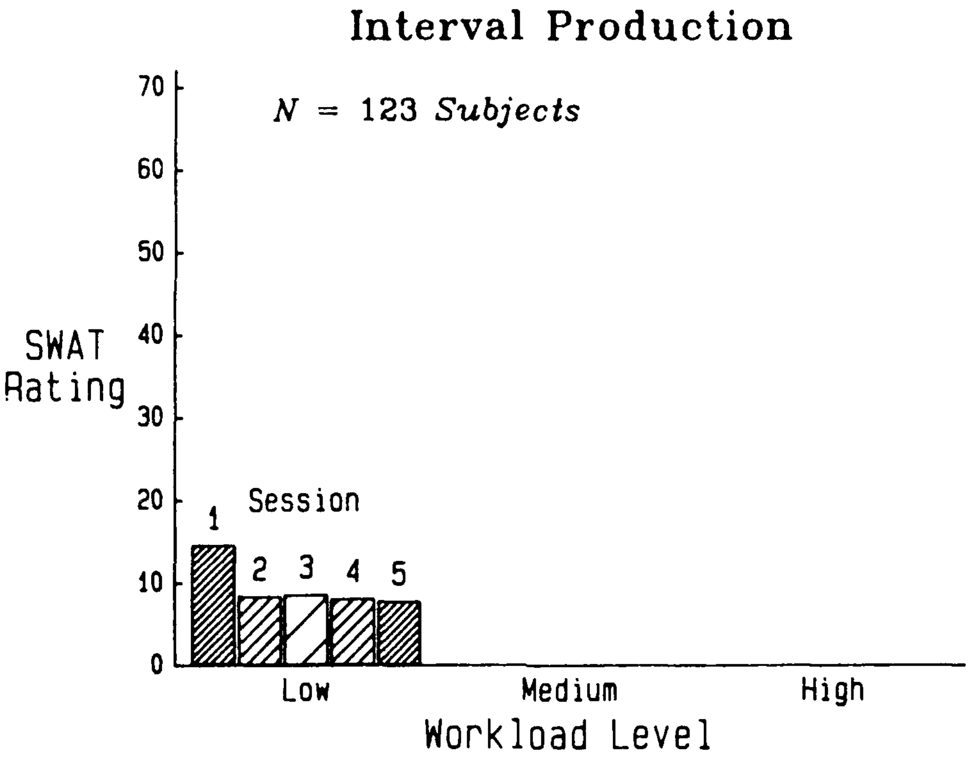


Figure 5-24. Mean SWAT Ratings for Interval Production - Trials 1 through 5.

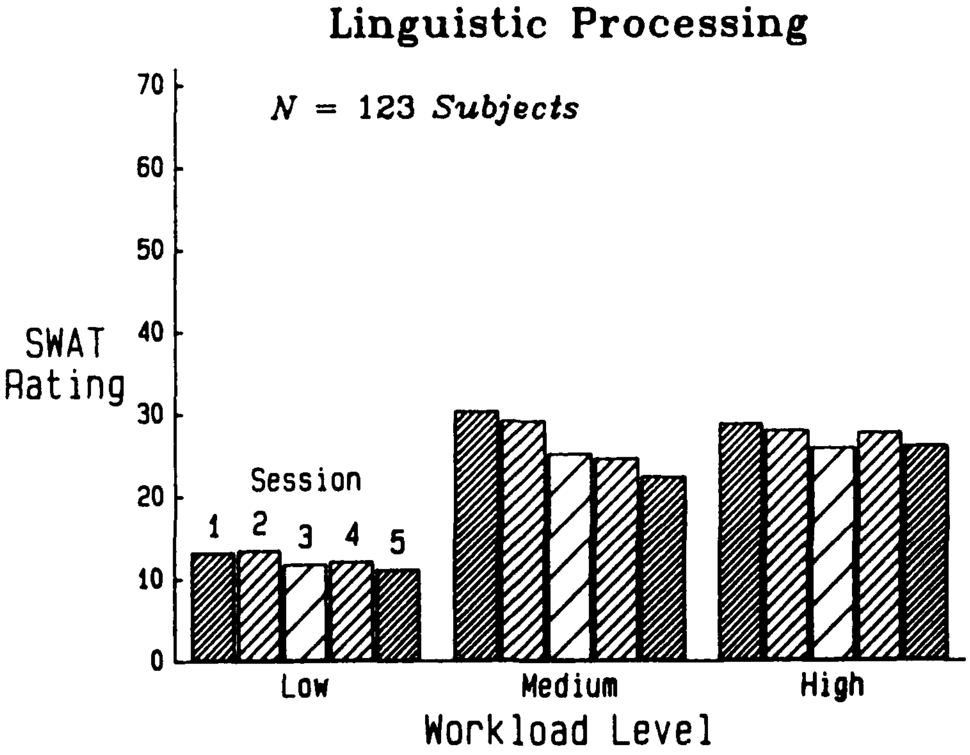


Figure 5-25. Mean SWAT Ratings for Linguistic Processing - Trials 1 through 5.

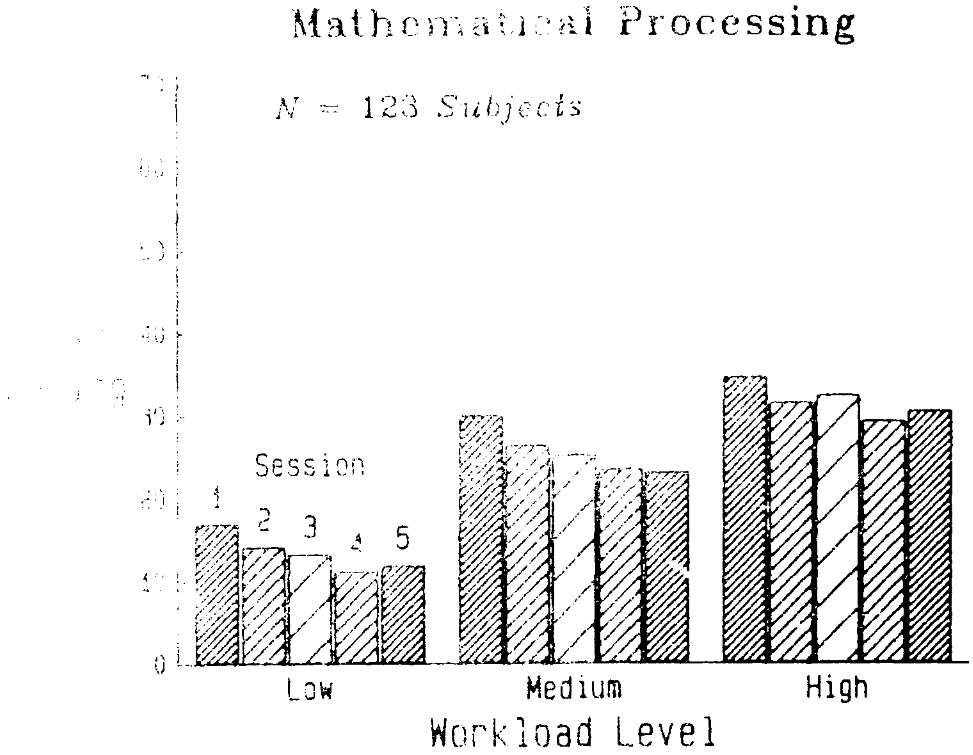


Figure 5-26. Mean SWAT Ratings for Mathematical Processing - Trials 1 through 5.

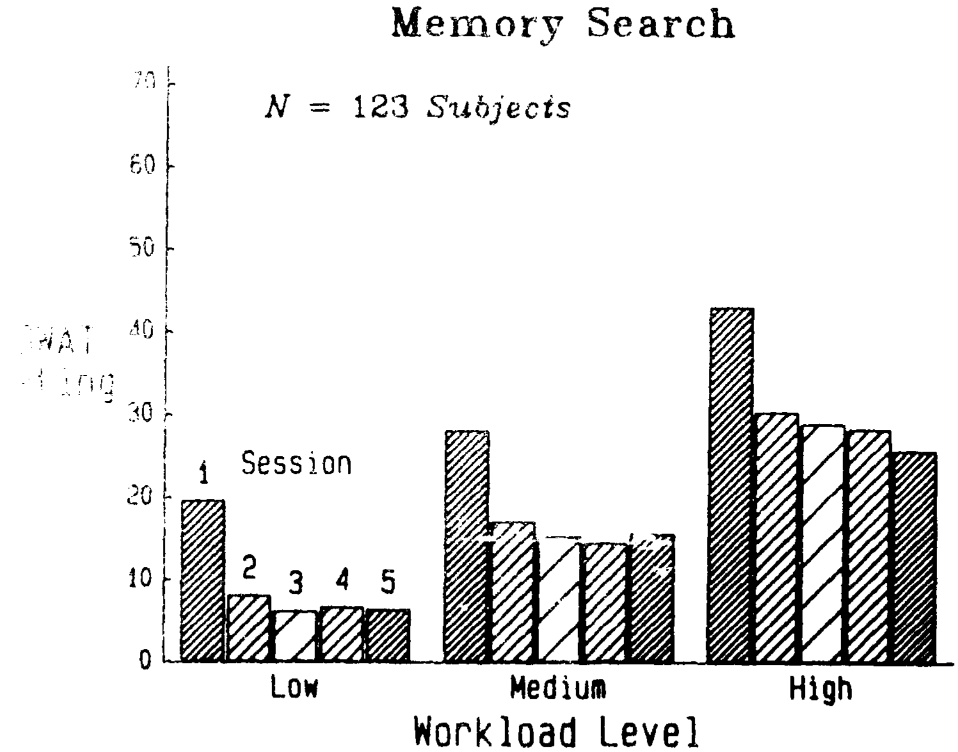
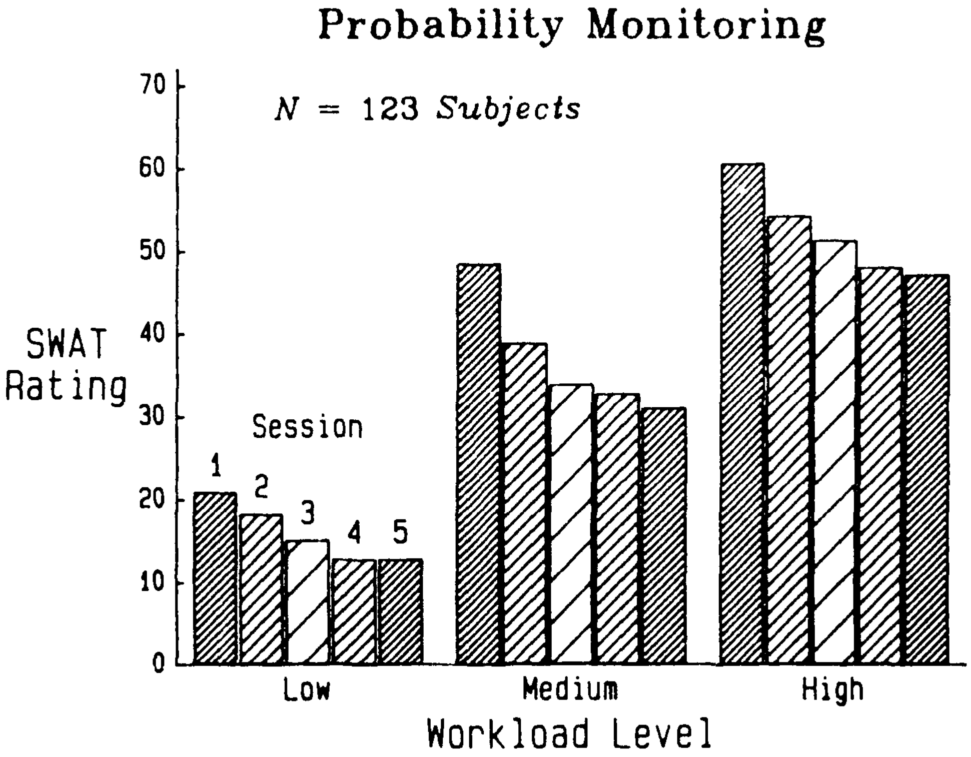


Figure 5-27. Mean SWAT Ratings for Memory Search - Trials 1 through 5.



F:gure 5-28. Mean SWAT Ratings for Probability Monitoring - Trials 1 through 5.

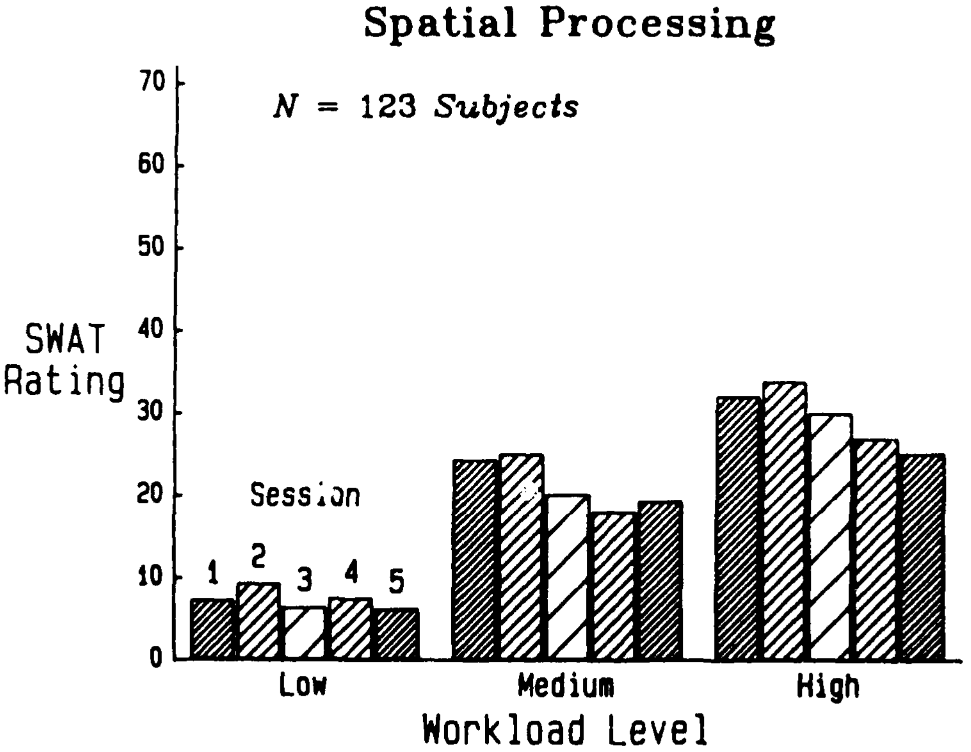


Figure 5-29. Mean SWAT Ratings for Spatial Processing - Trials 1 through 5.

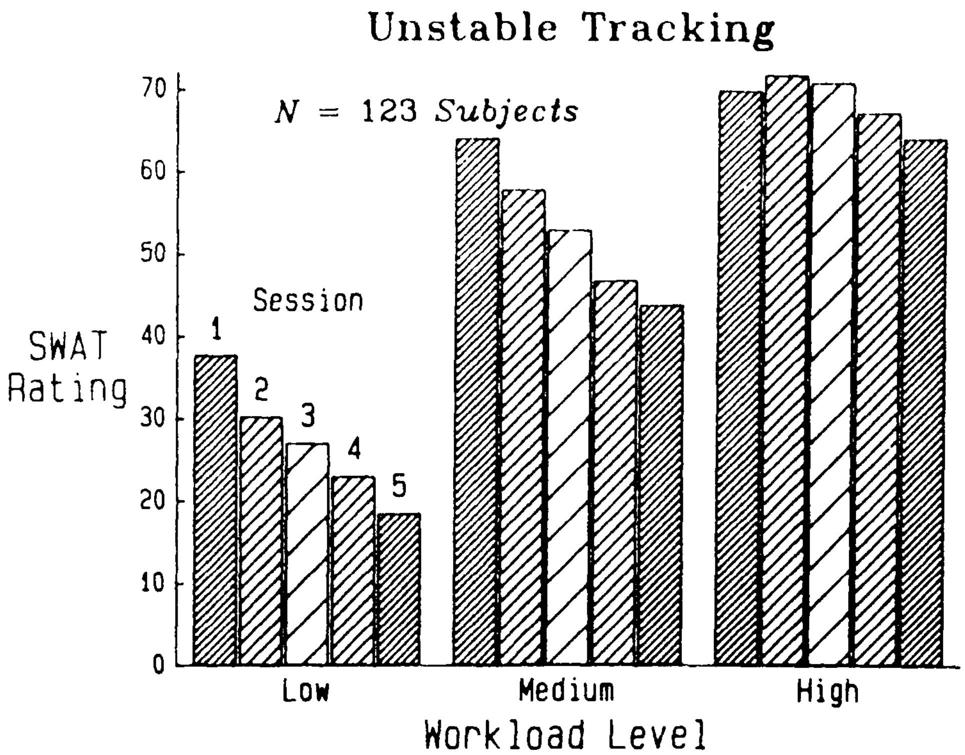


Figure 5-30. Mean SWAT Ratings for Unstable Tracking - Trials 1 through 5.

Table 5-21. ANOVA Summary for Level and Trial Effects by Task - Training Trials.

| Task | Model | Level | | Trial | | Level * Trial | |
|---|---|---|---|---|---|---|---|
| | $R^2$ | $F_{(2,244)}$ | $p>F$ | $F_{(4,488)}$ | $p>F$ | $F_{(8,976)}$ | $p>F$ |
| CR | 0.88 | 512.33 | * | 62.82 | * | 9.59 | * |
| GR | 0.87 | 217.24 | * | 8.04 | * | 3.66 | .0003 |
| IP | 1.00 | . | . | 8.54 | * | . | . |
| LP | 0.83 | 133.86 | * | 5.06 | .0005 | 2.00 | .0431 |
| MP | 0.87 | 224.26 | * | 7.26 | * | 0.55 | - |
| MS | 0.87 | 316.93 | * | 44.42 | * | 1.23 | - |
| PM | 0.90 | 464.26 | * | 24.42 | * | 3.39 | .0008 |
| SP | 0.85 | 303.37 | * | 9.76 | * | 3.36 | .0008 |
| UT | 0.91 | 425.57 | * | 25.99 | * | 8.70 | * |

* $p < 0.0001$

**Table 5-22. Significant Trial Differences by Task and Level.**

| Task | Level | $F_{(4,488)}$ | Trial | | | | |
|---|---|---|---|---|---|---|---|
| CR | L | 48.46 | 1 | 2 | 3 | 4 | 5 |
| | M | 66.18 | 1 | 2 | 3 | 4 | 5 |
| | H | 11.07 | 1 | 2 | 3 | 5 | 4 |
| GR | L | 14.54 | 1 | 2 | 3 | 4 | 5 |
| | M | 1.75 | 1 | 2 | 3 | 4 | 5 |
| | H | 3.41 | 2 | 1 | 4 | 3 | 5 |
| IP | - | 8.54 | 1 | 3 | 2 | 4 | 5 |
| LP | L | 1.01 | 2 | 1 | 4 | 3 | 5 |
| | M | 6.66 | 1 | 2 | 3 | 4 | 5 |
| | H | 1.26 | 1 | 2 | 4 | 5 | 3 |
| MP | L | 4.48 | 1 | 2 | 3 | 5 | 4 |
| | M | 6.23 | 1 | 2 | 3 | 4 | 5 |
| | H | 2.88 | 1 | 3 | 2 | 5 | 4 |
| MS | L | 36.45 | 1 | 2 | 4 | 5 | 3 |
| | M | 21.49 | 1 | 2 | 5 | 3 | 4 |
| | H | 26.89 | 1 | 2 | 3 | 4 | 5 |
| PM | L | 8.48 | 1 | 2 | 3 | 5 | 4 |
| | M | 24.67 | 1 | 2 | 3 | 4 | 5 |
| | H | 13.63 | 1 | 2 | 3 | 4 | 5 |
| SP | L | 1.66 | 2 | 4 | 1 | 3 | 5 |
| | M | 7.78 | 2 | 1 | 3 | 5 | 4 |
| | H | 9.02 | 2 | 1 | 3 | 4 | 5 |
| UT | L | 25.78 | 1 | 2 | 3 | 4 | 5 |
| | M | 29.73 | 1 | 2 | 3 | 4 | 5 |
| | H | 3.88 | 2 | 3 | 1 | 4 | 5 |

## 5.9 Summary of SWAT Results

Results from the scale development phase of the Subjective Workload Assessment Technique illustrated the varying ability of subjects to perform the SWAT sort. This emphasizes the importance of employing a relatively large sample size in studies involving subjective workload assessment. The majority (58%) of subjects were members of the Stress prototype group. The percentage of subjects in this group was much higher for females (78%, total *n* = 28) than for males (55%, total *n* = 95). Eighteen percent (18%) of the subjects were placed in the Time group, 13% in the Effort group and 11% in none of the groups. The last category contained subjects who identified equally with two or more prototypes or who provided extremely poor SWAT sorts.

Separate scaling solutions for the Time, Effort, and Stress prototype groups were developed for males and females. The Kendall coefficients of concordance verified a high level of agreement among subjects within each of the subgroups.

There was substantial variability in the range of SWAT values used by individual subjects in assessing the 25 task-level combinations. Some subjects used the full 0 to 100 range. Other subjects concentrated their ratings in a narrower range usually at the low or high end of the scale. The emphasis of a particular subject did not appear related to gender or prototype differences. These results point out the importance of establishing scale anchor points to assist subjects in calibrating their personal scales. This is particularly important when employing subjects with vastly differing experience reference points (fighter pilots vs. college freshmen).

For the baseline trials, the results were in general agreement with the performance data. With the exception of the Spatial Processing task, there were no differences between the SWAT ratings for the two baseline trials. Only minor differences existed between men and women with men providing lower ratings on spatial and input/output tasks and higher ratings on memory tasks. The Effort prototype gave generally higher ratings across all tasks.

The SWAT ratings validated the difficulty manipulation for all levels of all tasks. However, the distinction between the medium and high levels of Linguistic Processing, while significant, was slight in comparison with all other CTS tasks. This was also evident in the training data where the difference between these levels was not significant. The magnitude of the level effect ranged from 10% to 35% of the total variance. The magnitude of the subject effect ranged from 33% to 56%. The combined level and subject effects accounted for 52% to 70% of the total variance depending on the task. The highest accountable percentage of variance was associated with the Effort dimension for the level effect and with the Time dimension for the subject effect. In other words, Effort was the primary dimension used to distinguish difficulty level while Time was the dimension with the greatest subject variability.

Although the SWAT ratings in the current study verified the distinct workload levels for each task, there was little agreement with the ratings obtained in the Shingledecker (1984) parametric studies, perhaps due to the larger subject sample and exposure to a wider range of workload conditions. Cluster analysis of the SWAT ratings for the 25 task-level combinations yielded some overlap among the tasks with only a single distinct cluster (Unstable Tracking).

A final summary of the means and standard deviations of the SWAT ratings for all 123 subjects and both baseline trials is provided in Table 5-23.

**Table 5-23. Means (Standard Deviations) of SWAT Ratings by Task and Level for Trials 6 and 8 Combined (N = 123 Subjects).**

| Task | Level | | | | | |
|---|---|---|---|---|---|---|
| | Low | | Med | | High | |
| | $\bar{x}$ | $(s)$ | $\bar{x}$ | $(s)$ | $\bar{x}$ | $(s)$ |
| CR | 18.3 | (17.3) | 34.4 | (18.4) | 56.0 | (24.2) |
| GR | 27.3 | (20.5) | 36.5 | (22.5) | 50.1 | (25.9) |
| IP | 7.4 | (14.2) | . | . | . | . |
| LP | 11.7 | (16.0) | 22.8 | (19.5) | 27.5 | (20.0) |
| MP | 12.0 | (16.0) | 23.0 | (18.9) | 30.6 | (21.2) |
| MS | 7.7 | (14.4) | 14.8 | (17.2) | 26.7 | (21.1) |
| PM | 10.8 | (17.1) | 31.3 | (20.1) | 47.9 | (25.7) |
| SP | 7.2 | (14.7) | 18.4 | (17.6) | 24.8 | (19.9) |
| UT | 19.5 | (20.2) | 43.2 | (25.7) | 62.4 | (2[illegible]) |

# REFERENCES

Barratt, E. (1965). Factor Analysis of Some Psychometric Measures of Impulsiveness and Anxiety. *Psychological Reports, 16,* 547-554.

Broadbent, D. (1979). Human Performance and Noise. In C. Harris (Ed.), *Handbook of Noise Control.* New York: McGraw Hill.

Dodd, D.H. and Schultz, R.F. Jr. (1973). Computational Procedures for Estimating Magnitude of Effect for Some Analysis of Variance Designs. *Psychological Bulletin, 79,* 391-395.

Eschenbrenner, A.J., Jr. (1971). Effects of Intermittent Noise on the Performance of a Complex Psychomotor Task. *Human Factors, 13* (1), 59-63.

Eysenck, H.J. and Eysenck, S.B.G. (1968). *Manual for the Eysenck Personality Inventory.* Educational and Industrial Testing Service, San Diego, California.

Hockey, G. (1978). Effects of Noise on Human Work Efficiency. In D. May (Ed.), *Handbook of Noise Assessment.* New York: Van Nostrand Reinhold.

Jenkins, C., Zyzanski, S., and Rosenman, R. (1979). *Jenkins Activity Survey Manual.* The Psychological Corporation, New York, New York.

Kobasa. S.C. and Maddi, S.R. (1977). Existential Personality Theory. In R. Corsini (Ed.), *Current Personality Theories.* Itasca, Illinois: Peacock.

Mehrabian, A. (1976). *Manual for the Questionnaire Measure of Stimulus Screening.* Unpublished manuscript. Available from Albert Mehrabian, Los Angeles, California.

Michon, J.A. (1964). A Note on the Measurement of Perceptual Motor Load. *Ergonomics, 7,* 461-463.

Michon, J.A. (1966). Tapping Regularity as a Measure of Perceptual Motor Load. *Ergonomics, 9,* 401-412.

Murphy, D.L., Belmaker, R.H., Buchsbaum, M., Martin, N.F., Ciaranello, W., and Wyatt, R.J. (1977). Biogenic Amine-Related Enzymes and Personality Variations in Normals. *Psychological Medicine, 7,* 149-157.

Naitoh, P. and Townsend, R.E. (1970). The Role of Sleep Deprivation Research in Human Factors. *Human Factors, 12* (6), 575-585.

Nygren, T.E. (1982, April). *Conjoint Measurement and Conjoint Scaling: A Users Guide* (Tech. Report AFAMRL-TR-82-22). Wright-Patterson Air Force Base, Ohio: Air Force Aerospace Medical Research Laboratory.

Pasnau, R.O., Naitoh, P., Stier, S., and Kollar, E.J. (1968). The Psychological Effects of 205 Hours of Sleep Deprivation. *Archives of General Psychiatry, 18* , 496-505.

Reid, G.B. (1982). Subjective Workload Assessment: A Conjoint Scaling Approach. In *Aerospace Medical Association Annual Scientific Meeting* (pp. 153-154). Bal Harbor, Florida.

Reid, G.B., Eggemeier, F.T., and Nygren, T.E. (1982). An Individual Differences Approach to SWAT Scale Development. In *Proceedings of the Human Factors Society 26th Annual Meeting* (pp. 639-642). Seattle, Washington: Human Factors Society.

Reid, G.B., Shingledecker, C.A., and Eggemeier, F.T. (1981). Application of Conjoint Measurement to Workload Scale Development. In *Proceedings of the Human Factors Society 25th Annual Meeting* (pp. 522-526). Rochester, New York: Human Factors Society.

Sarason, I.G. (1972). Experimental Approaches to Test Anxiety: Attention and the Uses of Information. In Spielberger, C.D. (Ed.), *Anxiety: Current Trends in Theory and Research,* (Vol. II). New York: Academic Press.

*SAS User's Guide: Basics, Version 5 Edition* (1985). Cary, North Carolina: SAS Institute Inc.

*SAS User's Guide: Statistics, Version 5 Edition* (1985). Cary, North Carolina: SAS Institute Inc.

Schlegel, R.E. (1986). *Development of an Optimal Testing Protocol for the USAF Criterion Task Set (CTS)* (Final Scientific Report for contract SCEEE-84 RIP 47). University of Oklahoma, Norman, Oklahoma.

Schlegel, R.E. and Shingledecker, C.A. (1985). Training Characteristics of the Criterion Task Set Workload Assessment Battery. In *Proceedings of the Human Factors Society 29th Annual Meeting* (pp. 770-773). Baltimore, Maryland: Human Factors Society.

Shingledecker, C.A. (1984, November). *A Task Battery for Applied Human Performance Assessment Research* (Tech. Report AFAMRL-TR-84-071). Wright-Patterson Air Force Base, Ohio: Air Force Aerospace Medical Research Laboratory.

Shingledecker, C.A., Acton, W.H., and Crabtree, M.S. (1983, October). *Development and Application of a Criterion Task Set for Workload Metric Evaluation* (SAE Technical Paper 831419). Aerospace Congress & Exposition, Long Beach, California.

Spielberger, C., Gorsuch, R., and Lushene, R. (1970). *Manual for the State-Trait Anxiety Inventory.* Consulting Psychologist Press, Palo Alto, California.

Sternberg, S. (1969). Scanning Mental Processes Revealed by Reaction Time Experiments. *American Scientist. 57,* 421-457.

Thackray, R.I. (1982). Some Effects of Noise on Monitoring Performance and Physiological Response. *Academic Psychology Bulletin, 4,* 73-81.

Vaughan. G.M. and Corballis, M.C. (1969). Beyond Test of Significance: Estimating Strength of Effects in Selected ANOVA Designs. *Psychological Bulletin, 72,* 204-213.

Vidulich, M.A. and Tsang, P.S. (1986). Techniques of Subjective Workload Assessment: A Comparison of SWAT and the NASA-Bipolar Methods. *Ergonomics, 29,* 1385-1398.

Wickens, C.D. (1981). *Processing Resources in Attention, Dual Task Performance, and Workload Assessment* (Tech. Report EPL-81-3). University of Illinois, Illinois: Engineering Psychology Research Laboratory.

Wilkinson, R.T. (1964). Effect of Up to 60 Hours of Sleep Deprivation on Different Types of Work. *Ergonomics, 7,* 175-186.

Wilkinson, R.T. (1968). Sleep Deprivation. *Triangle, 8,* 162-166.

Williams, H.L., Lubin, A., and Goodnow, J.J. (1959). Impaired Performance with Acute Sleep Loss. *Psychological Monographs, 73* (14, Whole No. 484).

Winer, B.J. (1971). *Statistical Principles in Experimental Design.* New York: McGraw Hill.

Zuckerman, M. (1972). Sensation Seeking and Habituation of the Electrodermal Orienting Response. *Psychophysiology, 9,* 267-268.

Zuckerman, M. (1979). *Sensation Seeking: Beyond the Optimal Level of Arousal.* Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Zuckerman, M., Kolin, E.A., Price, L., and Zoob, I. (1964). Development of a Sensation Seeking Scale. *Journal of Consulting Psychology, 28,* 477-482.

☆U.S. Government Printing Office: 1990 — 748-002/00246